SPECTRAL METHODS FOR INITIAL BOUNDARY VALUE PROBLEMS

A Thesis Submitted
In Partial Fulfilment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

By
ASHOTOSH KUMAR SINGH

to the
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
JULY, 1992

To

My Parents

MATH-1992-D-SIN-SPE

CERTIFICATE

This is to certify that the matter embodied in the thesis entitled "SPECTRAL METHODS FOR INITIAL BOUNDARY VALUE PROBLEMS" by Mr. Ashotosh Kumar Singh for the award of Degree of Doctor of Philosophy of the Indian Institute of Technology Kanpur is a record of bonafide research work carried out by him under my supervision and guidance. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

July 1992

(Pravir Dutt)
Assistant Professor
Department of Mathematics
Indian Institute of Technology Kanpur
INDIA

ACKNOWLEDGEMENTS

CONTENTS

SYNOPSIS

Finite difference and finite element methods have been in practice for solving differential equations and their convergence analysis has proved that accuracy of these methods is of finite order. In problem like numerical weather prediction, numerical simulations of turbulent flows and other problems where high accuracy is desired for complicated solutions, these methods do not give very reliable results. Spectral methods have recently emerged as a viable alternative to these methods and have performed spectacularly well on many problems (e.g. in fluid dynamics where large hydrodynamics codes are now regularly used to study turbulence and transition, in numerical weather prediction and in ocean dynamics). In this thesis an attempt has been made to solve initial boundary value problems using spectral methods.

Spectral methods are based on representing the solution as the truncated series of orthogonal polynomial in the spatial variable and, in principle, are infinite order accurate.

Current formulation of spectral methods for solving initial boundary value problems employ a spectral discretization only in space and rely on finite difference techniques for advancing in time. As a result the global accuracy of the method is reduced to only finite order unless very small time steps are used which is not always practicable.

In the early 1980s Morchoisne [Rech. Aerosp. 1979-5, 293-306] proposed a method for solving such systems of equations

which was spectral in both space and time. However, even though his numerical results were impressive, the method has not acquired general acceptance so far. One reason for this is that it requires considerably more memory than conventional spectral methods. Another possible reason for the neglect of this approach is that it lacks a theoretical justification.

Recently, Dutt [Siam J. Numer. Anal., Vol 27, No 4, 885-903] proposed a method for solving initial boundary value problems which is similar to Morchoisne's approach in that it employs a spectral discretization in both space and time, henceforth we shall refer to it as the Galerkin-Collocation method. The method is set in a Galerkin formulation as it seeks an approximate solution which minimizes a weighted sum of the residuals in a filtered version of the partial differential equations and initial and boundary conditions.

The solution process, however, effectively amounts to collocating the filtered version of the partial differential equation and initial and boundary conditions at an overdetermined set of points. We show in this thesis that the filtering can be dispensed with, and it suffices to collocate the partial differential equation and initial and boundary conditions at the overdetermined set of points. The solution is then obtained by finding a least-squares solution to the overdetermined set of equations. It has been proved that the solution thus obtained converges to the actual solution at a spectral rate of accuracy in both space and time. In practice, the huge, full and overdetermined set of equations is solved by iterative techniques in which a low order finite difference solver is used as a

preconditioner.

Spectral methods give very accurate approximations to hyperbolic problems with smooth solutions. The naive use of spectral methods on hyperbolic problems with discontinuous solutions, however produces oscillatory numerical results. It has been known for some time that these oscillations are in themselves not insurmountable but contain sufficient information to permit reconstruction of the actual solution.

The observation that the pointwise convergence of a high order polynomial approximation to a discontinuous solution is very slow where as the convergence in a weighted mean is very fast suggests to post process the solution obtain by standard Collocation or Galerkin methods by a local smoothing in order to recover spectral accuaracy. Local smoothing will be carried out by a convolution in physical space with a localized function and hence by a weighted mean which approximates exceedingly well the exact value of the solution. From a mathematical point of view, the convergence in the mean can be measured in terms of Sobolev norm of negative order. It can be shown that the error between the computed and exact solution in a negative Sobolev norm decays at rate which depends only on the order of the norm.

In this thesis Galerkin-collocation method is used for solving hyperbolic initial boundary value problems and periodic initial value problems with nonsmooth data. The whole work is presented in the form of five chapters.

Chapter I is devoted to the introduction of Galerkin-Collocation method and a brief history of spectral methods is given.

Chapter II aims at the discussion and implementation of the Galerkin-collocation method to linear first order hyperbolic initial boundary value problems with variable coefficients in one space dimension using Chebyshev polynomials as trial functions. It is established that the filtering can be dispensed with and it suffices to collocate the partial differential equation and initial and boundary conditions at the over determined set of points.

In chapter III the system of equations obtained using spectral method discussed in chapter II is solved using iterative techniques. The preconditioned residual minimization method is discussed for scalar and for system case. This chapter also deals with the numerical treatment of boundary coditions which is essential for effective spectral calculations. Finally, in last section of the chapter we show how nonlinear problems can be solved spectrally using preconditioning. The efficacy of preconditioning is shown by computational results.

The objective of chapter IV is to discuss the Galerkin – Collocation method for solving hyperbolic initial boundary value problems using Legendre Polynomials as trial functions. Numerical results are given for scalar problems.

Finally, in chapter V we show that if we filter the data and solve periodic initial value problems with nonsmooth data using Galerkin Collocation method, then we can recover pointwise values with spectral accuracy, provided that the actual solution is piecewise smooth.

CHAPTER — I

GENERAL INTRODUCTION

1.1 Introduction

Many important equations of mathematical physics are
hyperbolic in nature e.g. Euler equations of gas dynamics,
Maxwell's equations, equations of Magneto hydrodynamics , the
classical and elastic wave equations. There is a constant demand
for efficient algorithms for solving these equations from
physicists and engineers. In this thesis a recently proposed
approach Dutt(1990) is used for solving hyperbolic initial
boundary value problems numerically which is spectral in
nature and this method is hereafter called as "Galerkin —
Collocation method" .

Before discussing Galerkin — Collocation method we need to
say a few words about spectral methods. Spectrals method may be
viewed as an extreme development of the class of discretization
schemes for differential equations known generically as the
method of weighted residual (MWR), (Finlayson and Seriven 1966).
The key elements of the (MWR) are the trial functions (also
called approximating functions) and test functions (also known as
weight functions). The trial functions are used as basis
functions for a truncated series expansion of the solution. The
test functions are used to ensure that the differential equation
is satisfied as closely as possible by the truncated series
expansion. This is achieved by minimizing the residual i.e. the

error in the differential equation produced by using the truncated expansion instead of exact solution, with respect to a suitable norm.

The choice of test functions distinguishes between the three most commonly used spectral schemes, namely the Galerkin, Collocation, and Tau version. In the Galerkin approach , the test functions are the same as the trial functions. They are, therefore, infinitely smooth functions which individually satisfy the boundary conditions. The differential equation is enforced by requiring that the integral of the residual times each test function be zero. In the Collocation approach the test functions are translated Dirac delta functions centered at special, so-called collocation points. This approach requires the differential equation to be satisfied exactly at the collocation points. Spectral Tau methods are similar to Galerkin methods in the way that the differential equation is enforced. However none of the test functions need satisfy the boundary conditions. Hence, a supplementary set of equations is used to apply the boundary conditions.

The spectral techniques applied in practice involve discretization of the spatial variable spectrally and use of a finite element or finite difference scheme for marching in time, this makes the method over all only finite ordered accurate. In early 1980's Morchoisne(1979, 1984) proposed a method for solving initial boundary value problems (IBVP) which was spectral in both space and time. However, even though his numerical results were impressive, the method has not acquired general acceptance so

far. One reason for this was that it required considerably more memory than conventional spectral method. Another possible reason for the neglect of this approach is it lacked a theoretical justification.

Recently Dutt(1990) gave an alternative approach for solving initial boundary value problems which is intermediate between Galerkin and Collocation methods and we refer to it as the Galerkin - Collocation method. Galerkin - Collocation method treats discretization in a different fashion and achieves spectral accuracy in both space and time. Dutt has provided a rigorous frame - work for Galerkin - Collocation method. For hyperbolic initial boundary value problems, Dutt has proved that the method is stable whenever the IBVP is stable and converges to the actual solution at a spectral rate of accuracy in both space and time.

Galerkin-Collocation method involves collocating the partial differential equation (PDE) and the initial and boundary conditions at an overdetermined set of collocation points. The approximate solution is a function, belonging to an appropriate finite dimensional space, which minimizes a weighted average of the residuals at these points. The finite dimensional space is usually the space of some Geggenbaur polynomials and the weights arise naturally from the Kreiss-Rauch estimate and the Gauss-Lobatto quadrature rule for the polynomials chosen. We explain how the method applies to a particularly simple hyperbolic initial boundary value problem

(1.1.1a) $\qquad u_t + u_x = 0 \qquad\qquad -1 \le x \le 1 \ , \ t \ge 0 \ ,$

with boundary condition

(1.1.1b) $u(-1,t) = g(t)$ $t \geq 0$,

and initial condition

(1.1.1c) $u(x,0) = f(x)$ $-1 \leq x \leq 1$.

Suppose $P_i(x)$ denotes a particular kind of Geggenbaur polynomial of degree i in x. To find out an approximation u^N to the solution of problem (1.1.1) in the space generated by $\{ P_i(x) \}_{0 \leq i \leq N}$ and $\{ P_j(t) \}_{0 \leq j \leq N}$. Then u^N can be written as

$$u^N(x,t) = \sum_{j=0}^{N} \sum_{i=0}^{N} a_{ij} P_i(x) P_j(t) .$$

We want to determine $\{ a_{ij} \}_{0 \leq i, j \leq N}$ such that u^N approximates the solution of (1.1.1) spectrally. To do this we do the following.

Let $\{ x_i \}_{0 \leq i \leq N}$ and $\{ t_j \}_{0 \leq j \leq N}$ denote the Gauss-Lobatto quadrature points for the polynomials we have chosen. We will collocate the PDE at $\{ (x_i, t_j) \}_{0 \leq i, j \leq N}$.

Suppose f^N and g^N are the unique polynomials of degree N such that they interpolate the functions f and g at $\{ x_i \}_{0 \leq i \leq N}$ and $\{ t_j \}_{0 \leq j \leq N}$ respectively.

$$f^N(x_i) = f(x_i) \qquad\qquad 0 \leq i \leq N$$

$$g^N(t_j) = g(t_j) \qquad\qquad 0 \leq j \leq N$$

Define the residuals

$$\rho_{ij} = (u_t^N + u_x^N)(x_i, t_j) \qquad\qquad 0 \leq i, j \leq N$$

$$\sigma_i = u^N(x_i, 0) - f^N(x_i) \qquad\qquad 0 \leq i \leq N$$

$$\tau_j = u^N(-1, t_j) - g^N(t_j) \qquad\qquad 0 \leq j \leq N$$

Ideally we will like to choose $\{ a_{ij} \}_{0 \leq i, j \leq N}$ such that all these

residuals are zero. In that case we will have to solve an overdetermined system of linear equations with a huge and full matrix. Instead we minimize certain weighted averages of the residuals ρ_{ij}, σ_i, τ_j, the weights being determined by the energy estimates and the quadrature rule.

It has been proved that the minimization problem we intend to solve is equivalent to obtaining a least-squares solution to the system of equations formed by enforcing the PDE and the initial and boundary conditions at an overdetermined set of collocation points. Because of this we have to resort to iterative techniques to obtain the least-squares solution.

Spectral methods give very highly accurate approximations to hyperbolic problems with smooth solutions. The naive use of spectral methods on hyperbolic problems with discontinuous solutions, however, produces oscillatory numerical results. The oscillations arising directly from the discontinuity have a Gibbs-like high frequency character. It has been known for some time that these oscillations are in themselves not insurmountable but contain sufficient information to permit reconstruction of the actual solution. This is achieved by a filtering of the computed values.

A detailed examination of the effect of filtering for linear systems of hyperbolic equations with periodic boundary conditions and discontinuous initial data was made by Majda *et.al.*(1978). They showed that for problems in one space dimension, it was possible to achieve a convergence rate of infinite order by a proper filtering of the initial conditions and also by applying a

filtering during derivative evaluations. However, in two space dimensions this infinite order of accuracy can be obtained only in a domain which excludes the region of influence and this region spreads linearly with time. Moreover, it is not clear as to how to handle problems where there are discontinuities in the forcing function.

As opposed to global smoothing , one can post-process the solution obtained by standard Collocation or Galerkin methods by a local smoothing in order to recover spectral accuracy. The idea is based on the observation that while the pointwise convergence of a high order polynomial approximation to a discontinuous solution is very slow, the convergence in a weighted mean is very fast. Local smoothing will be carried out by a convolution in physical space with a localized function and hence by a weighted mean which approximates exceedingly well the exact values of the solution.

From a mathematical point of view , the convergence in the mean can be measured in terms of a Sobolev norm of negative order. It can be shown that the error between the computed and exact solution in a negative Sobolev norm decays at a rate which depends only on the order of the norm. The idea was originally developed by Abarbanel *et.al.* (1986) , Gottlieb *et.al.* (1985) and Mercier(1981). In their formulation the approximate solution is obtained by first solving a system of ordinary differential equations, arising from either the Galerkin or Collocation method and then post processing is applied to the computed solution of this semi-discrete system of equations. We are not aware as to

how this procedure would deal with problems in which there are discontinuities in the forcing function also , instead of just in the initial data. We show in this thesis that for hyperbolic problems with periodic boundary conditions, it is possible to recover pointwise values with spectral accuracy using Galerkin – Collocation method even when there are discontinuities in the initial data and forcing function, as long as the actual solution is piecewise smooth.

In this thesis, we restrict ourselves to hyperbolic equations. The theory will be developed for systems in one space dimension. The case of periodic initial value problem with nonsmooth data is considered in the last chapter. The problem to be addressed here is the linear first order Hyperbolic Initial Boundary Value Problem with variable coefficients in one space dimension.

Consider the hyperbolic system of equations

(1.1.2a) $\qquad Lu(x,t) = F(x,t) \qquad\qquad 0 \leq x \leq 1 , \ t \geq 0$

where L is the differential operator given by

$$Lu = u_t - Au_x - Bu$$

$u(x,t), \ F(x,t)$ are vector valued functions with n components. $A(x,t), \ B(x,t)$ are (nxn) matrix valued functions in which each entry as a function of x and t is smooth.

The boundary conditions are given by

(1.1.2b) $\qquad Mu(0,t) = g(t) \qquad\qquad t \geq 0$

(1.1.2c) $\qquad Pu(1,t) = h(t) \qquad\qquad t \geq 0$

M and P are lxn and kxn matrix valued functions which specify the

l and k inflow variables in terms of the (n-1) and (n-k) outflow variables at the respective boundaries. Each entry in them as a function of t is smooth. g and h are vector valued functions of t with l and k components respectively.

The initial condition is given by

(1.1.2d) $u(x,0) = f(x)$ $0 \leq x \leq 1$

F, f, g and h are smooth and f, g and h are compatible at the space-time corner.

Since L is hyperbolic A has real eigenvalues and a complete set of real eigenvectors. In fact, A is similar to a diagonal matrix and without loss of generality we can assume that A itself is diagonal. We further assume that the boundary is noncharacteristic i.e. zero is not an eigenvalue of A at the boundary. Henceforth we assume that A can be written in the form

$$A = \begin{bmatrix} A^I & 0 \\ 0 & A^{II} \end{bmatrix} \qquad \begin{array}{l} A^I_{1 \times 1} < -\delta I_{1 \times 1} \\ A^{II}_{(n-1) \times (n-1)} > \delta I_{(n-1) \times (n-1)} \end{array}$$

where A^I and A^{II} are both diagonal and $\delta > 0$ for $x \in [0,1]$, $t \geq 0$

Kreiss established sufficient condition for problem (1.1.2) to be well posed. His condition is known as Uniform Kreiss Condition. Rauch modified it further to obtain a priori energy estimate for this type of problems. He proved that the following estimate holds.

If u is a solution of (1.1.2)

$$\left(\int_0^T \int_0^1 \| u(x,t) \|^2 \, dx dt \right)^{1/2} + \left(\int_0^1 \| u(x,T) \|^2 \, dx \right)^{1/2}$$

$$+ \left(\int_0^T \| u(0,t) \|^2 \, dt \right)^{1/2} + \left(\int_0^T \| u(1,t) \|^2 \, dt \right)^{1/2}$$

$$\leq C \, e^{-\eta t} \left[\int_0^T \int_0^1 \| F(x,t) \|^2 \, dxdt + \int_0^1 \| f(x) \|^2 \, dx \right.$$

$$\left. + \int_0^T \| g(t) \|^2 \, dt + \int_0^T \| h(t) \|^2 \, dt \right]^{1/2}$$

for all $\eta > \eta_0$, η_0 large enough and $T > 0$

The constant C is independent of T, F, f, g, h and η.

This is known as the Kreiss-Rauch estimate. Using this the Galerkin-Collocation method is developed for hyperbolic initial boundary value problems.

1.2 Layout of the thesis

The objective of the present thesis is to implement the Galerkin – Collocation method to hyperbolic problems with different trial functions and to show that the computed solution converges to the actual solution spectrally.

The work embodied in this thesis is divided into five chapters.

Chapter I is concerned with a brief history of spectral methods and introduction to Galerkin – Collocation method.

Chapter II aims at the discussion and the implementation of the Galerkin – Collocation method for solving initial boundary value problems. The trial functions are the Chebyshev polynomials. We show the filtering can be dispensed with and it suffices to collocate the partial differential equation and

initial and boundary condition at the over determined set of points.

Chapter III deals with preconditioning for scalar and system cases and nonlinear hyperbolic problems. We give the numerical treatment of boundary condition which is essential for effective spectral calculation. Computational results are presented to show the efficacy of the preconditioning at the end of each section.

In chapter IV we show how the Galerkin - Collocation method applies to hyperbolic initial boundary value problems using Legendre polynomials as trial functions. Numerical results are given for scalar problems.

Finally in chapter V we show that if we filter the data and solve periodic initial value problems with nonsmooth data using the Galerkin - Collocation method, then we can recover pointwise values with spectral accuracy, provided that the actual solution is piecewise smooth.

Galerkin – Collocation Method

2.1 Introduction

In this chapter we implement a spectral method for solving initial boundary value problems which is in between the Galerkin and Collocation Methods. In this method the partial differential equation and initial and boundary conditions are collocated at an overdetermined set of points and the approximate solution is chosen to be the least – squares solution to this system of equations. In the second section we have discussed the method and theoretical results.

2.2 Discussion of Method and Theoretical Results

In this chapter we restrict ourselves to the case of one space dimension. The method we describe, however, is applicable to any number of space dimensions.

We shall shift our initial time from $t = 0$ to $t = -1$ as this will considerably simplify our presentation. Consider the differential operator

$$(2.2.1) \qquad\qquad Lu \equiv u_t - Au_x - Bu.$$

Here u is a vector-valued function with k components and A and B are k × k matrix-valued functions which are smooth functions of x and t. We assume the system (2.2.1) is strictly hyperbolic. We consider the initial boundary value problem

$$(2.2.2a) \qquad L\, u(x,t) = F(x,t) \qquad , \text{for } -1 \le x \le 1 , -1 \le t \le 1,$$

with boundary conditions

(2.2.2b) M u(-1,t) = g(t)   ,for $-1 \leq t \leq 1$, and

(2.2.2c) P u(1,t) = h(t)   ,for $-1 \leq t \leq 1$,

and initial condition

(2.2.2d)   u(x,-1) = f(x)   ,for $-1 \leq x \leq 1$.

If there are ℓ inflow variables at the boundary x = -1 then M is a $\ell \times$ k matrix valued function which prescribes the ℓ inflow variables in terms of the (k-ℓ) outflow variables. Similarly if there are s inflow variables at the boundary x = 1 then P is a s $\times$ k matrix – valued function. Both M and P are smooth functions of t. We assume that the initial and boundary data f, g, h and forcing function F are smooth and satisfy the compatibility conditions which must hold at the space-time corners for the solution u to be smooth.   Finally we assume that the above initial boundary value problem (IBVP) satisfies the Uniform Kreiss condition.   If the Uniform Kreiss condition is satisfied then the IBVP is well posed, i.e. the solution u depends continuously on its data. More precisely, it has been proved that the following estimate

$$\int_{-1}^{1} \int_{-1}^{1} \| u(x,t) \|^2 dx \, dt + \int_{-1}^{1} \| u(-1,t) \|^2 dt + \int_{-1}^{1} \| u(1,t) \|^2 dt$$

(2.2.3) $$+ \int_{-1}^{1} \| u(x,1) \|^2 dx$$

$$\leq C \left[\int_{-1}^{1} \int_{-1}^{1} \| F(x,t) \|^2 dx \, dt + \int_{-1}^{1} \| f(x) \|^2 dx + \int_{-1}^{1} \| g(t) \|^2 dt \right.$$

$$+ \int_{-1}^{1} \| h(t) \|^2 dt \Bigg]$$

holds, for some positive constant C. Here the norm $\| \ \|$ denotes the Euclidean norm.

One final remark we make is that an IBVP that is well posed is 'structurally stable', i.e. if we perturb the coefficients of the differential operator and boundary operator by a small amount then the perturbed problem continues to remain well posed. This property is crucial for proving that the approximate solution we obtain by our method converges to the actual solution of the IBVP.

The method which we now describe applies to general Geggenbauer polynomials but in this chapter we shall describe it for Chebyshev polynomials. We recall that the Chebyshev polynomials $T_j(y) = \cos(j \cos^{-1}(y))$ are orthogonal with respect to the weight function

$$\omega(y) = \frac{1}{\sqrt{1-y^2}}$$

in the interval $[-1,1]$.

Let $S^{p,q}$ be the set of polynomials $w^{p,q}(x,t)$ of the form

(2.2.4) $$w^{p,q}(x,t) = \sum_{i=0}^{p} \sum_{j=0}^{q} a_{ij} T_i(x) T_j(t) ,$$

with scalar coefficients a_{ij}. Similarly, we shall denote by $(S^{p,q})^k$ the set of polynomials $w^{p,q}$ of the form (2.2.4) if the coefficients a_{ij} are vectors with k components. Henceforth we shall assume that there exists a constant λ such that $\frac{1}{\lambda} \le \frac{p}{q} \le \lambda$.

We now define an interpolation operator $I^{p,q}$ which takes a continuous function $r(x,t)$ defined on $[-1,1] \times [-1,1]$ and projects it into $S^{p,q}$. Thus

$$(2.2.5) \qquad I^{p,q} \, r(x,t) = \sum_{j=0}^{q} \sum_{i=0}^{p} b_{ij} \, T_i(x) \, T_j(t) = \overline{r}^{p,q}(x,t)$$

is the unique polynomial belonging to $S^{p,q}$ which interpolates $r(x,t)$ at the $(p+1) \times (q+1)$ points $\{(x_i^p, t_j^q)\}_{i=0,\ldots p, \ j=0,\ldots q}$.
Here the points

$$x_i^p = \cos(i\pi/p) \quad , \quad 0 \le i \le p, \quad \text{and}$$

$$t_j^q = \cos(j\pi/q) \quad , \quad 0 \le j \le q,$$

are the Gauss-Lobatto-Chebyshev points.

In much the same way we can define a one-dimensional interpolation operator I^{ℓ} which takes a continuous function $s(y)$ defined on $[-1,1]$ and projects it into the space of polynomials of degree $\le \ell$. Thus

$$(2.2.6) \qquad I^{\ell} s(y) = \sum_{i=0}^{\ell} b_i \, T_i(y) = \overline{s}^{\ell}(y)$$

is the unique polynomial of degree $\le \ell$ which interpolates $s(y)$ at the $(\ell+1)$ points $\{y_i^{\ell} = \cos(i\pi/\ell)\}_{i=0,\ldots\ell}$.
We can now use these interpolation operators to define a filtered version of the differential operator

$$L \, u = u_t - A \, u_x - Bu \, .$$

Let

$$\overline{A}^{p,q} = I^{p,q} A \quad , \quad \text{and}$$

$$\overline{B}^{p,q} = I^{p,q} B$$

be the polynomial interpolants of the $k \times k$ matrix-valued

functions A and B. We now define the differential operator

(2.2.7) $\qquad L^{p,q} u = u_t - \bar{A}^{p,q} u_x - \bar{B}^{p,q} u$,

which can be regarded as a perturbed version of the original differential operator Lu.

Similarly, we define

$$M^q = I^q M \quad , \quad \text{and}$$

$$P^q = I^q P \quad .$$

We now replace the original IBVP by a filtered version :

(2.2.8.a) $\qquad L^{p,q} \tilde{u}(x,t) = F(x,t)$, for $-1 \le x$, $t \le 1$,

with boundary conditions

(2.2.8.b) $\qquad M^q \tilde{u}(-1,t) = g(t)$, for $-1 \le t \le 1$,

(2.2.8.c) $\qquad P^q \tilde{u}(1,t) = h(t)$, for $-1 \le t \le 1$,

and initial conditions

(2.2.8.d) $\qquad \tilde{u}(x,-1) = f(x)$, for $-1 \le x \le 1$.

The above IBVP will be well posed if we choose p and q large enough. In fact since (2.2.8) can be regarded as a perturbation of (2.2.2) the following energy estimate, (Dutt 1990)

$$\int_{-1}^{1} \int_{-1}^{1} \| u^{p,q}(x,t) \|^2 dx \, dt + \int_{-1}^{1} \| u^{p,q}(-1,t) \|^2 dt$$

$$+ \int_{-1}^{1} \| u^{p,q}(1,t) \|^2 dt + \int_{-1}^{1} \| u^{p,q}(x,1) \|^2 dx$$

$$\le C \left[\int_{-1}^{1} \int_{-1}^{1} \| L^{p,q} u^{p,q}(x,t) \|^2 dx \, dt + \int_{-1}^{1} \| u^{p,q}(x,-1) \|^2 dx \right.$$

(2.2.9) $\qquad \left. + \int_{-1}^{1} \| M^q u^{p,q}(-1,t) \|^2 dt + \int_{-1}^{1} \| P^q u^{p,q}(1,t) \|^2 dt \right],$

holds for p and q large enough , with some constant C. Henceforth we shall let C denote a generic constant.

From (2.2.9) the inequality

$$\int_{-1}^{1}\int_{-1}^{1}||u^{p,q}(x,t)||^2 dxdt \leq C \left[\int_{-1}^{1}\int_{-1}^{1}||L^{p,q}u^{p,q}(x,t)||^2\omega(x)\omega(t)dxdt \right.$$

$$+ \int_{-1}^{1}||u^{p,q}(x,-1)||^2\omega(x)dx + \int_{-1}^{1}||M^q u^{p,q}(-1,t)||^2\omega(t)dt$$

$$(2.2.10) \qquad + \left. \int_{-1}^{1}||P^q u^{p,q}(1,t)||^2\omega(t)dt \right] ,$$

follows immediately since the weight function $\omega \geq 1$.

We wish to find an approximate solution $u^{p,q}(x,t) \in (S^{p,q})^k$ to the above IBVP. Notice that if $u^{p,q}(x,t) \in (S^{p,q})^k$ then

$$L^{p,q} u^{p,q}(x,t) \in (S^{2p,2q})^k ,$$

$$M^q u^{p,q}(-1,t) \in (S^{2q})^{\ell} ,$$

$$P^q u^{p,q}(1,t) \in (S^{2q})^s , \text{ and}$$

$$u^{p,q}(x,-1) \in (S^p)^k ,$$

and this suggests that we should accordingly filter our data. Let

$$\overline{F}^{2p,2q}(x,t) = I^{2p,2q} F(x,t) ,$$

$$\overline{g}^{2q}(x,t) = I^{2q} g(t) ,$$

$$\overline{h}^{2q}(t) = I^{2q} h(t) , \text{ and}$$

$$\overline{f}^{2p}(x) = I^{2p} f ,$$

be filtered representations of the data. If we were to substitute our approximate solution into the IBVP the residuals

$$\rho^{p,q}(x,t) = L^{p,q} u^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t) ,$$

$$\sigma^q(t) \quad = \quad M^q \, u^{p,q}(-1,t) - \overline{g}^{2q}(t) \quad ,$$

(2.1.11)
$$\eta^q(t) \quad = \quad P^q \, u^{p,q}(1,t) - \overline{h}^{2q}(t) \quad ,$$

$$\tau^p(x) \quad = \quad u^{p,q}(x,-1) - \overline{f}^{2p}(x) \quad ,$$

would, in general not be zero. We would like to choose our
approximate solution $u^{p,q}(x,t)$ so that it makes these residuals
as small as possible and for this we need to define a functional
which will measure the size of the residuals.

Accordingly we define a functional

$$H^{p,q}(v^{p,q}) = \int_{-1}^{1} \int_{-1}^{1} ||L^{p,q} \, v^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t)||^2 \omega(x)\omega(t) \; dxdt$$

$$+ \int_{-1}^{1} ||M^q v^{p,q}(-1,t) - \overline{g}^{2q}(t)||^2 \omega(t) dt$$

$$+ \int_{-1}^{1} ||P^q v^{p,q}(1,t) - \overline{h}^{2q}(t)||^2 \omega(t) dt$$

(2.2.12)
$$+ \int_{-1}^{1} || v^{p,q}(x,-1) - \overline{f}^{2p}(x) ||^2 \omega(x) dx,$$

where

$$v^{p,q}(x,t) = \sum_{i=0}^{p} \sum_{j=0}^{q} b_{ij} \, T_i(x) \, T_j(t) \in (\, S^{p,q})^k \; .$$

We choose as our approximate solution the unique $u^{p,q} \in (S^{p,q})^k$
which minimizes a functional $H^{p,q}(v^{p,q})$ over all $v^{p,q}$, where
$H^{p,q}(v^{p,q})$ is essentially equivalent to $H^{p,q}(v^{p,q})$. Now we
observe that

$$\rho^{p,q}(x,t) \quad = \quad L^{p,q} \, v^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t) \in (\, S^{2p,2q})^k,$$

$$\sigma^q(t) \quad = \quad M^q \, v^{p,q}(-1,t) - \overline{g}^{2q}(t) \in (S^{2q})^\ell \; ,$$

$$\eta^q(t) \quad = \quad P^q \, v^{p,q}(1,t) - \overline{h}^{2q}(t) \in (S^{2q})^s \; , \text{ and}$$

$$\tau^p(x) \quad = \quad v^{p,q}(x,-1) - \overline{f}^{2p}(x) \in (S^{2p})^k \quad ,$$

and so we can exactly evaluate the integrals in (2.2.12) by using the very highly accurate Gauss quadrature rules. In particular, for the Gauss-Lobatto-Chebyshev rule we have that if s(y) is a polynomial of degree $\leq 2N - 1$ then

$$(2.2.13) \qquad \int_{-1}^{1} s(y)\, \omega(y) dy \; = \; \frac{\pi}{N} \sum_{j=0}^{N} \frac{s(y_j^N)}{c_j^N} \; ,$$

where the points y_j^N are given by

$$y_j^N \; = \; \cos(\pi j/N) \; , \qquad 0 \leq j \leq N,$$

and the weights c_j^N are given by

$$c_j^N \; = \; 2 \qquad \text{if} \quad j \neq 0 \quad \text{or} \quad N,$$

$$c_j^N \; = \; 1 \qquad \text{otherwise.}$$

However, there is a stronger version of this rule which we use for our particular case. Suppose r(y) is a polynomial of degree $\leq N$. Then the inequality ,(Canuto *et.al.* 1987)

$$(2.2.14) \qquad \int_{-1}^{1} r^2(y)\omega(y) dy \; \leq \; \frac{\pi}{N} \sum_{j=0}^{N} \frac{r^2(y_j^N)}{c_j^N} \; \leq \; 2 \int_{1}^{1} r^2(y)\, \omega(y) \, dy$$

holds.

We can therefore replace the functional $H^{p,q}(v^{p,q})$ we are trying to minimize by an equivalent functional

$$
\begin{aligned}
H^{p,q}(v^{p,q}) = {} & \frac{\pi^2}{4pq} \sum_{j=0}^{2q} \sum_{i=0}^{2p} \frac{\| \, L^{p,q}\, v^{p,q}(x_i^{2p}, t_j^{2q}) - \overline{F}^{2p,2q}(x_i^{2p}, t_j^{2q}) \, \|^2}{c_i^{2p} \times c_j^{2q}} \\[2mm]
& + \frac{\pi}{2q} \sum_{j=0}^{2q} \frac{\| \, M^{q}\, v^{p,q}(-1, t_j^{2q}) - \overline{g}^{2q}(t_j^{2q}) \, \|^2}{c_j^{2q}} \\[2mm]
& + \frac{\pi}{2q} \sum_{j=0}^{2q} \frac{\| \, P^{q}\, v^{p,q}(1, t_j^{2q}) - \overline{h}^{2q}(t_j^{2q}) \, \|^2}{c_j^{2q}}
\end{aligned}
$$

(2.2.15)

$$+ \frac{\pi}{2p} \sum_{i=0}^{2p} \frac{\| v^{p,q}(x_i^{2p}, -1) - \overline{f}^{2p}(x_i^{2p}) \|^2}{c_i^{2p}}$$

In fact, using (2.2.14) we conclude that

$$\overline{H}^{p,q}(v^{p,q}) \leq H^{p,q}(v^{p,q}) \leq 4 \overline{H}^{p,q}(v^{p,q}).$$

We choose as our approximate solution $u^{p,q} \in (S^{p,q})^k$ which minimizes $H^{p,q}$.

In other words, our solution $u^{p,q}$ is given by a least-squares solution to the overdetermined system of equations

$$\left(\frac{\pi^2}{4p\, q\, c_i^{2p}\, c_j^{2q}} \right)^{1/2} \left\{ (L^{p,q} u^{p,q} - F)(x_i^{2p}, t_j^{2q}) \right\} = 0,$$

$$0 \leq i \leq 2p \ , \ 0 \leq j \leq 2q,$$

$$\left(\frac{\pi}{2q\, c_j^{2q}} \right)^{1/2} \left\{ M^q(t_j^{2q})\, u^{p,q}(-1, t_j^{2q}) - g(t_j^{2q}) \right\} = 0,$$

(2.2.16)
$$0 \leq j \leq 2q,$$

$$\left(\frac{\pi}{2q\, c_j^{2q}} \right)^{1/2} \left\{ P^q(t_j^{2q})\, u^{p,q}(-1, t_j^{2q}) - h(t_j^{2q}) \right\} = 0,$$

$$0 \leq j \leq 2q,$$

$$\left(\frac{\pi}{2p\, c_i^{2p}} \right)^{1/2} \left\{ u^{p,q}(x_i^{2p}, -1) - f)(x_i^{2p}) \right\} = 0,$$

$$0 \leq i \leq 2p.$$

Here, we have used the fact that $\overline{F}^{2p,2q}(x_i^{2p}, t_j^{2q}) = F(x_i^{2p}, t_j^{2q})$ etc. and so can work with point values of the original data. We may write the system (2.2.16) in the form

(2.2.17)
$$D^{p,q} U^{p,q} = Z^{p,q},$$

where $D^{p,q}$ is a $\lambda \times \nu$ matrix, $U^{p,q}$ is a ν column vector formed by concatenating the point values $\left\{ U^{p,q}(x_i^p, t_j^q) \right\}_{\substack{i=0,\ldots,p \\ j=0,\ldots,q}}$ and $Z^{p,q}$ is a λ-column vector with

$\lambda = k \times (2p+1) \times (2q+1) + (\ell+s) \times (2q+1) + k \times (2p+1)$, and

$\nu = k \times (p+1) \times (q+1)$.

We emphasize that $U^{p,q}$ denotes the ν column vector defined above and $u^{p,q}(x,t)$ the polynomial belonging to $(S^{p,q})^k$ whose point values are the components of $U^{p,q}$. We wish to find a least squares solution to the problem (2.2.17). Clearly, $U^{p,q}$ must satisfy the linear system of equations

$$\left[(D^{p,q})^T D^{p,q} \right] U^{p,q} = (D^{p,q})^T Z^{p,q}.$$

In (Dutt 1990) it has been shown that the matrix $(D^{p,q})^T (D^{p,q})$ has an inverse for p and q large enough. Hence the solution to the minimization problem is unique.

To store the filtered representations of the coefficient matrices $A^{p,q}, B^{p,q}$ etc. would place a prohibitive overhead on memory requirements for realistic problems; there is a way of getting around this, however. Instead of solving the system (2.2.17) we choose $\tilde{U}^{p,q}$ which is the least squares solution to the unfiltered system of equations

$$\left(\frac{\pi^2}{4(p\, q\, c_i^{2p}\, c_j^{2q})} \right)^{1/2} \left\{ L\tilde{u}^{p,q}(x_i^{2p}, t_j^{2q}) - F(x_i^{2p}, t_j^{2q}) \right\} = 0,$$

$$0 \le i \le 2p \ , \ 0 \le j \le 2q,$$

$$\left(\frac{\pi}{2q\, c_j^{2q}} \right)^{1/2} \left\{ M(t_j^{2q})\, \tilde{u}^{p,q}(-1, t_j^{2q}) - g(t_j^{2q}) \right\} = 0 \ ,$$

(2.2.18) $\qquad\qquad\qquad\qquad\qquad\qquad\qquad 0 \le j \le 2q,$

$$\left(\frac{\pi}{2q\, c_j^{2q}} \right)^{1/2} \left\{ P(t_j^{2q})\, \tilde{u}^{p,q}(1, t_j^{2q}) - h(t_j^{2q}) \right\} = 0 \ ,$$

$$0 \le j \le 2q,$$

$$\left(\frac{\pi}{2p \ c_j^{2q}}\right)^{1/2} \left\{\tilde{u}^{p,q}(x_i^{2p},-1) - f(x_i^{2p})\right\} = 0 \ ,$$

$$0 \leq i \leq 2p,$$

as our approximate solution.

Notice that the system (2.2.18) may be written as

(2.2.19) $\qquad\qquad \tilde{D}^{p,q} \ \tilde{U}^{p,q} = Z^{p,q}$

which is the same as (2.2.17) except that the matrix $D^{p,q}$ has been replaced by $\tilde{D}^{p,q}$, where $\tilde{D}^{p,q}$ may be regarded as a perturbed version of $D^{p,q}$. If $A(x,t)$ is a smooth function then we know that

$$\left|A(x_i^{2p},t_j^{2q}) - \bar{A}^{p,q}(x_i^{2p},t_j^{2q})\right|$$

is spectrally small for all i and j. Using this we conclude that the matrix $\tilde{D}^{p,q}$ differs from $D^{p,q}$ by a spectrally small amount and hence the difference between $U^{p,q}$ and $\tilde{U}^{p,q}$, the least -squares solutions of (2.2.17) and (2.2.19) respectively, is spectrally small.

We make the argument we have outlined above rigorous in the following lemmas and theorem.

Lemma 2.2.1

Let $v^{p,q}$ belong to the space of polynomials $S^{p,q}$ defined in (2.2.4), Then

(2.2.20) $\qquad \left(\int_{-1}^{1} \int_{-1}^{1}(v^{p,q})^2 \ \omega(x)\omega(t)dxdt\right) \leq C_1\left(\int_{-1}^{1} \int_{-1}^{1}(v^{p,q})^2 \ dxdt\right)$

where $C_1 = E_\alpha \ (pq)^{2-2/\alpha}$, for any $\alpha > 2$. In particular, choosing $\alpha = 4$ we obtain $C_1 = E_4 (pq)^{3/2}$

Using Holder's inequality we get

$$\left(\int_{-1}^{1} \int_{-1}^{1} (v^{p,q})^2 \, \omega(x)\omega(t) \, dx \, dt \right) \leq$$

$$\left(\int_{-1}^{1} \int_{-1}^{1} \left((v^{p,q})^2 \right)^{\alpha} dx \, dt \right)^{1/\alpha} \left(\int_{-1}^{1} \int_{-1}^{1} \left(\omega(x)\omega(t) \right)^{\beta} dx \, dt \right)^{1/\beta}$$

where $1/\alpha + 1/\beta = 1$. Now

$$\int_{-1}^{1} \left(\omega(x) \right)^{\beta} dx = \int_{-1}^{1} \frac{1}{(\; 1-x^2 \;)^{\beta/2}} \, dx$$

is finite for $1 < \beta < 2$.

For a fixed value of β the right hand side of the last equation becomes a constant which we shall denote by D_{β}. Hence we can conclude that

$$\left(\int_{-1}^{1} \int_{-1}^{1} \left(\omega(x)\omega(t) \right)^{\beta} dx \, dt \right)^{1/\beta} = D_{\beta}^{2/\beta}.$$

Now if $s(y)$ is a polynomial of degree m then by Nikolskii's inequality (Canuto *et.al.* 1987, p 288)

$$\left(\int_{-1}^{1} \left(s(y) \right)^{\alpha} dy \right)^{1/\alpha} \leq Km^{2(1/\beta - 1/\alpha)} \left(\int_{-1}^{1} \left(s(y) \right)^{\beta} dy \right)^{1/\beta},$$

for $1 \leq \beta < \alpha$.

Thus we obtain

$$\left(\int_{-1}^{1} \int_{-1}^{1} \left((v^{p,q})^2 \right)^{\alpha} dx dt \right)^{1/\alpha} \leq K^2 (4pq)^{2(1-1/\alpha)} \left(\int_{-1}^{1} \int_{-1}^{1} \left((v^{p,q})^2 \right) dx dt \right)$$

Putting $E_{\alpha} = K^2 (4D_{\beta})^{2/\beta}$ we get the required result. ∎

Lemma 2.2.2

There are constants K_1 and K_2 depending on p and q such that the estimate

(2.2.21)
$$K_1 \parallel V^{p,q} \parallel^2 \leq \parallel D^{p,q}V^{p,q} \parallel^2 \leq K_2 \parallel V^{p,q} \parallel^2 .$$

holds. Here

$$K_1 = C/\, p^5 \quad , \text{ and}$$

$$K_2 = C\, p^2 \quad , \text{ where C denotes a generic constant.}$$

Let $v^{p,q}(x,t) = \displaystyle\sum_{j=0}^{q} \sum_{i=0}^{p} a_{i,j} T_i(x)\, T_j(t)$ be the polynomial $\in (S^{p,q})^k$

such that $v^{p,q}(x_i^p, t_j^q) = \{V^{p,q}\}_{i,j}$ for $i=0,\ldots,p$, $j=0,\ldots,q$.

We have that

$$\parallel D^{p,q}V^{p,q} \parallel^2 = \frac{\pi^2}{4pq} \sum_{j=0}^{2q} \sum_{i=0}^{2p} \frac{\parallel L^{p,q}\, v^{p,q}(\, x_i^{2p}, t_j^{2q}\,)\parallel^2}{c_i^{2p}\, c_j^{2q}}$$

$$+ \frac{\pi}{2q} \sum_{j=0}^{2q} \frac{\parallel M^q\, v^{p,q}(-1, t_j^{2q})\parallel^2}{c_j^{2q}} + \frac{\pi}{2q} \sum_{j=0}^{2q} \frac{\parallel P^q\, v^{p,q}(1, t_j^{2q})\parallel^2}{c_j^{2q}}$$

$$+ \frac{\pi}{2p} \sum_{i=0}^{2p} \frac{\parallel v^{p,q}(x_i^{2p}, -1)\parallel^2}{c_i^{2p}} .$$

Then by (2.2.14) we obtain

$$\parallel D^{p,q}\, v^{p,q} \parallel^2 \leq 4 \left[\int_{-1}^{1} \int_{-1}^{1} \parallel L^{p,q}\, v^{p,q}(\, x,t\,)\parallel^2 \omega(x)\omega(t)dxdt \right.$$

$$+ \int_{-1}^{1} \parallel M^q\, v^{p,q}(-1,t)\parallel^2 \omega(t)\, dt + \int_{-1}^{1} \parallel P^q\, v^{p,q}(-1,t)\parallel^2 \omega(t)dt$$

$$\left. + \int_{-1}^{1} \parallel v^{p,q}(x,-1)\parallel^2 \omega(x)\, dx \right] .$$

Now by the inverse inequality for differentiation (Canuto *et.al.* 1987, p 295) if $v^{p,q} \in (S^{p,q})^k$ then

$$\int_{-1}^{1} \int_{-1}^{1} \| \frac{\partial}{\partial x} v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt$$

$$\leq Cp^4 \int_{-1}^{1} \int_{-1}^{1} \| v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt,$$

and

$$\int_{-1}^{1} \int_{-1}^{1} \| \frac{\partial}{\partial t} v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt$$

$$\leq Cq^4 \int_{-1}^{1} \int_{-1}^{1} \| v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt,$$

where by C we denote a generic constant.

Further, since $\sup_{(x,t)\in[-1,1]\times[-1,1]} \| A(x,t) \| \leq C$, and

$$\sup_{(x,t)\in[-1,1]\times[-1,1]} \| B(x,t) \| \leq C ,$$

we obtain

$$\int_{-1}^{1} \int_{-1}^{1} \| L^{p,q} v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt$$

$$\leq C(p^4+q^4) \int_{-1}^{1} \int_{-1}^{1} \| v^{p,q}(x,t) \|^2 \omega(x)\omega(t)dxdt$$

$$\leq C\frac{(p^4+q^4)}{pq} \sum_{j=0}^{q} \sum_{i=0}^{p} \| v^{p,q}(x_i^p, t_j^q) \|^2 ,$$

using (2.2.14).

Hence we can conclude that

$$\int_{-1}^{1} \int_{-1}^{1} \| L^{p,q} v^{p,q}(x,t) \|^2 \omega(x)\omega(t) \, dxdt \leq C\frac{(p^4+q^4)}{pq} \| v^{p,q} \|^2$$

Now

$$\int_{-1}^{1} ||v^{p,q}(x,-1)||^2 \, \omega(x) \, dx \le \frac{\pi}{q} \sum_{j=0}^{q} ||v^{p,q}(x^p,-1)||^2$$

$$\le \frac{Cp}{pq} \sum_{j=0}^{q} \sum_{i=0}^{p} ||v^{p,q}(x_i^p, t_j^q)||^2 \ .$$

Hence we can conclude that

$$\int_{-1}^{1} ||v^{p,q}(x,-1)||^2 \, \omega(x) dx \le \frac{Cp}{pq} || V^{p,q} ||^2,$$

and by similar arguments that

$$\int_{-1}^{1} || M^q \, v^{p,q}(-1,t)||^2 \, \omega(t) dt \le \frac{Cq}{qp} || V^{p,q} ||^2,$$

$$\int_{-1}^{1} || P^q \, v^{p,q}(1,t)||^2 \, \omega(t) dt \le \frac{Cq}{qp} || V^{p,q} ||^2.$$

Combining all the above inequalities we conclude that

$$|| D^{p,q} v^{p,q}||^2 \le C \left(\frac{p^4 + q^4}{pq} \right) || V^{p,q} ||^2 \ .$$

Now using the condition that there is a constant λ such that

$$\frac{1}{\lambda} \le \frac{p}{q} \le \lambda$$

we obtain $\quad || D^{p,q} \, v^{p,q} ||^2 \le Cp^2 || V^{p,q} ||^2.$

Next, we have to bound $|| D^{p,q} \, v^{p,q} ||^2$ from below . Using (2.2.14) we have that

$$||D^{p,q} v^{p,q}||^2 \ge \int_{-1}^{1} \int_{-1}^{1} ||L^{p,q} \, v^{p,q}(x,t)||^2 \omega(x)\omega(t) \, dxdt$$

$$+ \int_{-1}^{1} ||M^q \, v^{p,q}(-1,t)||^2 \omega(t) dt + \int_{-1}^{1} ||P^q \, v^{p,q}(1,t)||^2 \omega(t) dt.$$

$$+ \int_{-1}^{1} ||v^{p,q}(x,-1)||^2 \omega(x) dx.$$

And so by (2.2.10) we can conclude that

$$\|D^{p,q}V^{p,q}\|^2 \geq C \int_{-1}^{1} \int_{-1}^{1} \|v^{p,q}(x,t)\|^2 \, dx \, dt \ ,$$

and this together with lemma 2.2.1 and (2.2.14) gives us

$$\|D^{p,q}V^{p,q}\|^2 \geq \frac{C}{p^5} \|V^{p,q}\|^2$$

■

Theorem 2.2.1

The difference between $U^{p,q}$ and $\tilde{U}^{p,q}$, the solutions of equations (2.2.17) and (2.2.19) respectively, is spectrally small

Proof :

We have that

$$U^{p,q} = \{(D^{p,q})^T D^{p,q}\}^{-1} (D^{p,q})^T Z^{p,q} \ , \text{and}$$

$$\tilde{U}^{p,q} = \{(\tilde{D}^{p,q})^T \tilde{D}^{p,q}\}^{-1} (\tilde{D}^{p,q})^T Z^{p,q} .$$

Hence

$$\| U^{p,q} - \tilde{U}^{p,q} \|$$

(2.2.22)
$$\leq \|\{(D^{p,q})^T D^{p,q}\}^{-1}\| \ \|(D^{p,q})^T - (\tilde{D}^{p,q})^T\| \ \| Z^{p,q} \|$$

$$+ \|\{(\tilde{D}^{p,q})^T \tilde{D}^{p,q}\}^{-1} - \{(D^{p,q})^T D^{p,q}\}^{-1}\| \ \|(\tilde{D}^{p,q})^T\| \|Z^{p,q}\|.$$

Now we know that $\|\tilde{D}^{p,q} - D^{p,q}\| = O(\frac{1}{p^s})$ for any $s > 0$

And in lemma 2.1 we have shown that

$$K_1 \| V^{p,q} \|^2 \leq \| D^{p,q}V^{p,q} \|^2 \leq K_2 \| V^{p,q} \|^2 \ ,$$

which implies that

$$1/K_1 \geq \| \{(D^{p,q})^T D^{p,q}\}^{-1} \| \geq 1/K_2 \ .$$

Hence

(2.2.23)
$$\|\{(D^{p,q})^T D^{p,q}\}^{-1}\| \; \|(D^{p,q})^T - (\tilde{D}^{p,q})^T\| \; \|Z^{p,q}\|$$

$$\leq 1/K_1 O(\frac{1}{p^s})\|Z^{p,q}\|$$

We now estimate the second term in the R.H.S of (2.2.22).

Clearly

$$\| (\tilde{D}^{p,q}) \| \leq \| (D^{p,q}) \| + \| (\tilde{D}^{p,q}) - (D^{p,q}) \|$$

$$\leq 2\sqrt{K_2} + O(\frac{1}{p^s}) \; \leq 2\sqrt{K_2} \; , \text{ for } p,q \text{ large enough.}$$

Put

M = $(D^{p,q})^T D^{p,q}$, and

N = $\{(\tilde{D}^{p,q})^T \tilde{D}^{p,q}$.

Let

ΔM = $(\tilde{D}^{p,q})^T \tilde{D}^{p,q} - (D^{p,q})^T D^{p,q}$

Then N = M + ΔM. It is easy to show that $\| \Delta M \| = O(\frac{1}{p^s})$, for all s > 0 . Hence

$$\| N^{-1} - M^{-1}\| \leq \frac{\| M^{-1}\|^2 \| \Delta M \|}{1 - \| M^{-1}\| \, \| \Delta M \|} \; \leq \; 2 \| M^{-1}\|^2 \| \Delta M \|, \text{ for } p,q$$

large enough.

So we obtain

$$\| N^{-1} - M^{-1}\| \; \leq \frac{2}{K_1^2} O(\frac{1}{p^s}) \; .$$

Thus

(2.2.24)
$$\|\{(\tilde{D}^{p,q})^T \tilde{D}^{p,q}\}^{-1} - \{(D^{p,q})^T D^{p,q}\}^{-1}\| \; \| (\tilde{D}^{p,q})^T\| \; Z^{p,q} \|$$

$$\leq \quad 4 \frac{\sqrt{K_2}}{K_1^2} \; O(\frac{1}{p^s}) \| \; Z^{p,q} \; \|$$

And

$$\| Z^{p,q} \| \leq \sup_{(x,t)\in[-1,1]\times[-1,1]} \| F(x,t) \| + \sup_{t\in[-1,1]} \| g(t) \|$$

$$+ \sup_{t\in[-1,1]} \| h(t) \| + \sup_{x\in[-1,1]} \| f(x) \| \; .$$

Now combining (2.2.22), (2.2.23) and (2.2.24) we get the result

$$\| \; U^{p,q} - \widetilde{U}^{p,q} \; \| \leq \quad O(\frac{1}{p^s}) \; , \text{for all } s > 0.$$
∎

CHAPTER III

Iterative Solution Of Spectral Equations
For Chebyshev Polynomials

3.1   Introduction

The system of equations

(3.1.1) $L^{p,q} \ U^{p,q} = Z^{p,q}$

we get using spectral methods discussed in chapter II is huge,
full, ill - conditioned and overdetermined. To obtain a
least–squares solution to this problem we resort to iterative
techniques. In the first section of this chapter we describe the
numerical method for scalar problems. In the second section we
have discussed the method for the systems case and a numerical
treatment of boundary conditions is dealt with. Finally in
section 3 we have considered the method for nonlinear problem .

For notational convenience we drop superscript p,q in
$u^{p,q}(x,t)$, $U^{p,q}$ and $Z^{p,q}$ so that $u^{p,q}(x,t) \equiv u(x,t)$, $U^{p,q} \equiv U$ and
$Z^{p,q} \equiv Z$. Recall that

$$u^{p,q}(x,t) = \sum_{i=0}^{p} \sum_{j=0}^{q} a_{ij}^{p,q} \ T_i(x) \ T_j(t)$$

denotes the approximate solution of (3.1.1) and

$$U^{p,q} = \left\{ u^{p,q}(x_i^p, t_j^q) \right\}_{\substack{i=0;\dots;p \\ j=0;\dots;q}}$$

is the vector whose components are the $(p+1) \times (q+1)$ values of
$u^{p,q}$ evaluated at the Gauss-Chebyshev Lobatto points. We can then

write (3.1.1) in the equivalent form

(3.1.2) $L^{sp} U = Z$.

3.2 Preconditioning for scalar problems

A least-square solution V of the equation

$$L^{sp} U = Z ,$$

is obtained by minimizing the residual

$$H(V) = ||L^{sp}V - Z||^2 ,$$

and this suggests that we should seek the solution by using preconditioned residual minimization. For this we need to have an approximate inverse, which we shall denote by $(L^{ap})^{-1}$, to the matrix L^{sp} and we typically use a low order finite difference solver for $(L^{ap})^{-1}$. The method can then be described as :

1) Given the current guess $U^{(n)}$ compute the residual

$$R^{(n)} = Z - L^{sp} U^{(n)}$$

2) Obtain an improvement $V^{(n)}$ for $U^{(n)}$ by computing

$$V^{(n)} = (L^{ap})^{-1} R^{(n)}$$

3) Update the current value of $U^{(n)}$ by putting

$$U^{(n+1)} = U^{(n)} + \omega_n V^{(n)},$$

where ω_n is chosen as that value of ω at which $H(U^{(n)}+\omega V^{(n)})$ achieves its minimum. ω_n can be computed using the formula

$$\omega_n = \frac{\left(R^{(n)}, L^{sp} V^{(n)} \right)}{\left(L^{sp} V^{(n)}, L^{sp} V^{(n)} \right)} ,$$

where (,) denotes the standard inner product.

We now explain each step in more detail.

Given the $(p+1) \times (q+1)$ values of $U^{(n)}$ which are the point values of $u^{(n)}(x,t)$ at the points $\left\{ (x_i^p, t_j^q) \right\}_{\substack{i=0;\ldots;p \\ j=0;\ldots;q}}$ we compute $\Gamma^{(n)}$, the coefficients in its representation as a Chebyshev series

$$u^{(n)}(x,t) = \sum_{i=0}^{p} \sum_{j=0}^{q} \gamma_{ij}^{(n)} T_i(x) T_j(t).$$

This can be implemented using either a two dimensional Fast Chebyshev Transform or alternatively by matrix multiplications. As the details of this are well known (Orszag 1980 and Pulliam *et.al.* 1981) we do not go into it any further. Since we need to compute the residuals on a grid with $(2p+1) \times (2q+1)$ points we pad the representation of $u^{(n)}$ with zeros as

$$(3.2.1) \qquad u^{(n)}(x,t) = \sum_{i=0}^{2p} \sum_{j=0}^{2q} \gamma_{ij}^{(n)} T_i(x) T_j(t) \ ,$$

where $\gamma_{ij}^{(n)} = 0$ for $i>p$ or $j>q$. We can now calculate the values of $u^{(n)}(x,t)$ at the $(2p+1) \times (2q+1)$ points $\left\{ (x_i^{2p}, t_j^{2q}) \right\}_{\substack{i=0;\ldots;p \\ j=0;\ldots;q}}$ by using an inverse transform or matrix multiplications. It is now an easy matter tocompute the residuals

$$\rho_{ij}^{(n)} = \left(u_t^{(n)} - a\, u_x^{(n)} - b u^{(n)} - F \right)\left(x_i^{2p}, t_j^{2q} \right) \ ,$$

$$\sigma_{ij}^{(n)} = \left(M(t_j^{2q})\, u^{(n)}(-1, t_j^{2q}) - g(t_j^{2q}) \right) \qquad ,$$

$$(3.2.2)$$

$$\eta_j^{(n)} = \left(P(t_j^{2q})\, u^{(n)}(-1, t_j^{2q}) - h(t_j^{2q}) \right) \qquad ,$$

and $\qquad \tau_j^{(n)} = \left((u^{(n)}(x_i^{2p}, -1) - f(x_i^{2p}) \right) \ .$

For the scalar problem we shall denote the matrices A and B by a

and b.

The differentiations involved in computing (3.2.2) can be implemented using matrix multiplications or transform techniques. What is important to note is that these computations can be speeded up immensely using vectorization, as was pointed out by Orszag (1980). Henceforth we shall denote the vector of residuals $\left(\rho^{(n)},\ \sigma^{(n)},\ \eta^{(n)},\ \tau^{(n)}\right)$ by $R^{(n)}$.

2) We now seek an iterative improvement to the vector $U^{(n)}$ which we denote by the vector $V^{(n)} = \left\{ v^{(n)}(x_i{}^p, t_j{}^q) \right\}_{\substack{i=0,\ldots,p \\ j=0,\ldots,q}}$ with $(p+1) \times (q+1)$ components. Let $W^{(n)}$ denote the prolongation of $V^{(n)}$ onto the grid with $(2p+1) \times (2q+1)$ points $\left\{ (x_i^{2p}, t_j^{2q}) \right\}_{\substack{i=0,\ldots,p \\ j=0,\ldots,q}}$ as described in (3.2.1). We can write this as

$$(3.2.3) \qquad W^{(n)} = P\ V^{(n)}$$

where P denotes the prolongation operator.

It is natural to seek $V^{(n)}$ as the solution of the system of equations

$$(3.2.4) \qquad L^{ap}\ V^{(n)} = L^{fd}\ P\ V^{(n)} = R^{(n)}$$

where L^{fd} is a finite difference discretization of the IBVP on the finer mesh. Then we have

$$V^{(n)} = P^{-1} \left(L^{fd} \right)^{-1} R^{(n)}\ ,$$

where P^{-1} should be interpreted as the generalized inverse of the operator P. We can write this in two steps as

<div align="center">Compute</div>

a) $$W^{(n)} = \left(L^{fd} \right)^{-1} R^{(n)}$$

<div align="center">Compute</div>

b) $$V^{(n)} = P^{-1} W^{(n)} \ .$$

We describe these steps further.

a)   In computing $W^{(n)}$ it is important to choose the finite difference operator so that is easily invertible and stable. A first or second order implicit approximate factorization code based on central differencing ideally fulfills all these objectives (Pullaim *et.al.* 1981 & Steger *et.al.* 1985). In fact , many of the general purpose simulation codes in use in research and industry utilize just this approach and it should be possible to modify them to perform spectral calculations. Further, since these codes involve the solution of a set of independent tridiagonal or block – tridiagonal matrix solvers the solution process can be vectorized. We refer the interested reader to (Pullaim *et.al.* 1981 & Steger *et.al.* 1985) for details.

We indicate the equations obtained from the finite difference discretization of

(3.2.5) $$w_t^{(n)} - a(x,t) \, w_x^{(n)} - b(x,t) \, w^{(n)} = \rho^{(n)}(x,t)$$

at interior points of the space time square using implicit central differencing. Here $W^{(n)}$ denotes   the vector with $(2p+1) \times (2q+1)$ values $\left\{ w^{(n)}(x_i^{2p}, t_j^{2q}) = w_{ij}^{(n)} \right\}_{\substack{i=0;\ldots;2p \\ j=0;\ldots;2q}}$.
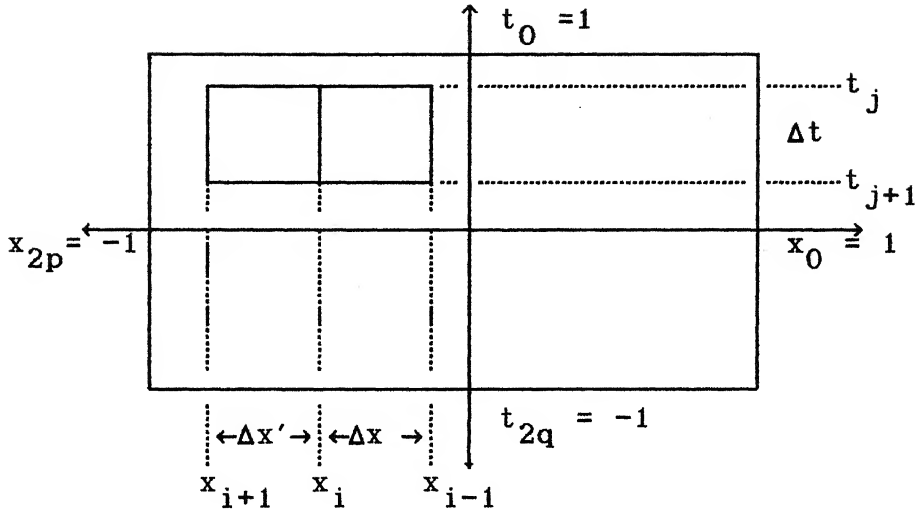
Fig 3.2.1

Let $a_{ij} = a(x_i^{2p}, t_j^{2q})$ and $b_{ij} = b(x_i^{2p}, t_j^{2q})$.

To advance the solution from time t_{j+1} to t_j we can use the implicit scheme

$$\frac{w_{i,j} - w_{i,j+1}}{\Delta t} - \frac{a_{ij}}{2}\left[\left\{\frac{w_{i-1,j+1} - w_{i,j+1}}{\Delta x} + \frac{w_{i,j+1} - w_{i+1,j+1}}{\Delta x'}\right.\right.$$

$$- \frac{w_{i-1,j+1} - w_{i+1,j+1}}{\Delta x' + \Delta x}\left\} + \left\{\frac{w_{i-1,j} - w_{i,j}}{\Delta x} + \frac{w_{i,j} - w_{i+1,j}}{\Delta x'}\right.\right.$$

$$\left.\left.- \frac{w_{i-1,j} - w_{i+1,j}}{\Delta x + \Delta x'}\right\}\right] - b_{ij}\left\{\frac{w_{i,j} + w_{i,j+1}}{2}\right\} = \rho_{i,j}$$

which is second order accurate.

This can be written as

$$\alpha_i\, w_{i-1,j} + \beta_i\, w_{i,j} + \gamma_i\, w_{i+1,j} = \rho_{i,j} - \left\{ \tilde{\alpha}_i\, w_{i-1,j+1} + \right.$$

(3.2.6)
$$\left. + \tilde{\beta}_i\, w_{i,j+1} + \tilde{\gamma}_i\, w_{i+1,j+1} \right\}$$

where

$$\alpha_i = \left\{ -\frac{1}{\Delta x} + \frac{1}{\Delta x + \Delta x'}\right\}\frac{a_{ij}}{2} \ , \ \tilde{\alpha}_i = \left\{ -\frac{1}{\Delta x} + \frac{1}{\Delta x' + \Delta x}\right\}\frac{a_{ij}}{2}$$

$$\beta_i = \left\{ -\frac{1}{\Delta t'} - \frac{a_{ij}}{2} \left(-\frac{1}{\Delta x} + \frac{1}{\Delta x'} \right) - \frac{b_{ij}}{2} \right\} ,$$

$$\tilde{\beta}_i = \left\{ -\frac{1}{\Delta t'} - \frac{a_{ij}}{2} \left(-\frac{1}{\Delta x} + \frac{1}{\Delta x'} \right) - \frac{b_{ij}}{2} \right\} ,$$

$$\gamma_i = \left\{ \frac{1}{\Delta x} + \frac{1}{\Delta x + \Delta x'} \right\} \frac{a_{ij}}{2} , \quad \tilde{\gamma}_i = \left\{ \frac{1}{\Delta x} - \frac{1}{\Delta x + \Delta x'} \right\} \frac{a_{ij}}{2} .$$

The above equation uses information from the 6 point stencil shown in Figure 3.2.1. Thus to advance from time level t_{j+1} to t_j we have to solve a tridiagonal system. To initialize the procedure we impose the initial conditions

(3.2.7) $\qquad\qquad w_{i,2q}^{(n)} = \tau_i^{(n)} , \qquad\qquad 0 \le i \le 2p.$

We can impose the boundary conditions either implicitly or explicitly. Inflow boundary conditions pose no problem. Thus if $x = -1$ is an inflow boundary for (3.2.5) we simply impose the boundary condition

(3.2.8) $\qquad\qquad w_{2p,j}^{(n)} = \sigma_j^{(n)} , \qquad\qquad 0 \le j \le 2q.$

If it is an outflow boundary, however, we either impose the partial differential equation at the boundary implicitly or use extrapolation techniques (Dutt 1990). Our computational results show that the implicit treatment is preferable and so we shall say a few words about it.
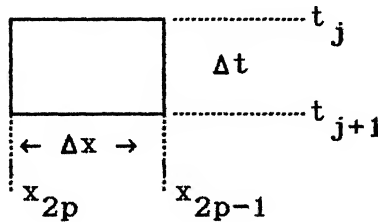

Fig 3.2.2

We use the four point stencil shown in Figure 3.2.2 to obtain the

equation

$$\left\{ \frac{w_{2p,j} - w_{2p,j+1}}{\Delta t} + \frac{w_{2p-1,j} - w_{2p-1,j+1}}{\Delta t} \right\}$$

$$- \frac{a_{2p,j}}{2} \left\{ \frac{w_{2p-1,j} - w_{2p,j}}{\Delta x} + \frac{w_{2p-1,j+1} - w_{2p,j+1}}{\Delta x} \right\}$$

$$- b_{2p,j} \left\{ \frac{w_{2p,j} + w_{2p,j+1} + w_{2p-1,j} + w_{2p-1,j+1}}{4} \right\} = \sigma_j.$$

This is of the form

(3.2.9) $\alpha_{2p} \, w_{2p-1,j} + \beta_{2p} \, w_{2p,j} = \sigma_j - \left\{ \tilde{\alpha}_{2p} \, w_{2p-1,j+1} + \tilde{\beta}_{2p} \, w_{2p,j+1} \right\}$,

where

$$\alpha_{2p} = \frac{1}{2\Delta t} - \frac{a_{2p,j}}{2\Delta x} - \frac{b_{2p,j}}{4} \,, \quad \tilde{\alpha}_{2p} = \frac{1}{2\Delta t} - \frac{a_{2p,j}}{2\Delta x} - \frac{b_{2p,j}}{4}$$

$$\beta_{2p} = - \frac{1}{2\Delta t} + \frac{a_{2p,j}}{2\Delta x} - \frac{b_{2p,j}}{4} \,, \quad \tilde{\beta}_{2p} = - \frac{1}{2\Delta t} + \frac{a_{2p,j}}{2\Delta x} - \frac{b_{2p,j}}{4}.$$

Note, that with this treatment of the boundary condition our system of equations remains tridiagonal, and with this we conclude our discussion of (a).

Computation for Scalar case

Example 3.2.1

$$u_t - (x + \epsilon(x,t)) \, u_x = F(x,t), \qquad -1 \le x \le 1, \ -1 \le t \le 1$$

$$F(x,t) = \epsilon(x,t) \times \sin(\tfrac{\pi}{16} x e^t) \times \tfrac{\pi}{16} e^t \,,$$

with initial conditions

$$u(x,-1) = \sin(\tfrac{\pi}{16} \, x e^{-1}) \,,$$

and boundary condition

$$u(-1,t) = \sin(\tfrac{\pi}{16} \, e^t) \text{ and } u(1,t) = \sin(\tfrac{\pi}{16} \, e^t).$$

The results obtained for three different values of $\epsilon(x)$.

1) $\epsilon_1(x,t) = 0.5 \times \sin(3\pi(x+t))$,

2) $\epsilon_2(x,t)=0.5\times\sin(4\pi(x+t))$,

3) $\epsilon_3(x,t)=0.5\times\sin(5\pi(x+t))$,

are shown below.

N - Number of collection points

N	$0.5\times\sin(3\pi(x+t))$	$0.5\times\sin(4\pi(x+t))$	$0.5\times\sin(5\pi(x+t))$
33	<2> (5.03×10^{-3}) <26> (2.69×10^{-10})	<2> (6.69×10^{-3}) <30> (2.32×10^{-10})	<3> (6.96×10^{-3}) <43> (3.75×10^{-10})

Table 3.2.1

The number inside the first bracket gives the iteration number at which the residual given in the second bracket is obtained.

Example 3.2.2

$$u_t + x\,u_x = 0$$

with initial condition

$$u(x,-1) = f(x)$$

Results obtained for three sets of initial data

(i) $f(x) = \sin(\frac{\pi}{16} x)$

(ii) $f(x) = \sin(\frac{\pi}{33} x)$

(iii) $f(x) = \sin(\frac{\pi}{100} x)$

are shown in table below. N - Number of collocation points

N	$\sin(\frac{\pi}{100} x)$	$\sin(\frac{\pi}{33} x)$	$\sin(\frac{\pi}{16} x)$
33	<2> (2.11×10^{-8}) <9> (5.01×10^{-14})	<2> (4.65×10^{-8}) <13> (8.58×10^{-14})	<2> (8.86×10^{-8}) <12> (1.75×10^{-12})

Table 3.2.2

Notice that in Example 3.2.1 u is an inflow variable at x = -1 and x = 1, that is why the value of u is prescribed at both the boundaries. In Example 3.2.2 u is an outflow variable at x = -1 and x = 1, and so the value of u at both the boundaries is obtained by enforcing the partial differential equation there. Orszag (1980) had advocated a filter in which the top one third of the frequency components of the numerical solution are removed and which he has refered to as the two-thirds rule, for the preconditioning to be effective. In table 3.2.3 and table 3.2.4 we, give comparative results for the two - thirds filter and the one - half filter, advocated by us.

$$u_t - a\ u_x = 0$$
$$u(x,-1) = f(x)$$

Number of collocation points is 33,

$$a = x$$

f(x)	1/2 filter	2/3 filter
$\sin\frac{\pi}{100}x$	$\langle 2\rangle\ (1.71\times10^{-6})$ $\langle 58\rangle(4.46\times10^{-14})$	$\langle 2\rangle\ (5.22\times10^{-6})$ $\langle 65\rangle(3.32\times10^{-7})$
$\sin\frac{\pi}{33}x$	$\langle 2\rangle\ (5.21\times10^{-6})$ $\langle 51)(8.67\times10^{-13})$	$\langle 2\rangle\ (1.55\times10^{-5})$ $\langle 60\rangle(8.07\times10^{-7})$
$\sin\frac{\pi}{16}x$	$\langle 3\rangle\ (6.60\times10^{-6})$ $\langle 50\rangle(2.81\times10^{-12})$	$\langle 2\rangle\ (3.03\times10^{-5})$ $\langle 50\rangle(1.75\times10^{-6})$

Table 3.2.3

$$a = -x$$

f(x)	1/2 filter	2/3 filter
$\sin\frac{\pi}{100}x$	<2> (2.11×10^{-8}) <9> (5.01×10^{-14})	<2> (2.09×10^{-8}) <10> (6.43×10^{-10})
$\sin\frac{\pi}{33}x$	<2> (4.65×10^{-8}) <13> (8.98×10^{-14})	<2> (4.57×10^{-8}) <16> (1.74×10^{-8})
$\sin\frac{\pi}{16}x$	<2> (8.86×10^{-8}) <12> (1.75×10^{-12})	<2> (1.16×10^{-7}) <15> (7.29×10^{-8})

Table 3.2.4

It can be seen from Table 3.2.3 and Table 3.2.4 that the one-half filter performs better than the two-thirds filter.

3.3 Preconditioning for System case

Consider the following hyperbolic system

$(3.3.1a)$ $w_t - A w_x - Bw = F$ $-1 \leq x, t \leq 1$

where

$$w = (w_1, w_2),$$

$$A = \begin{pmatrix} a_{11}(x,t) & a_{12}(x,t) \\ a_{21}(x,t) & a_{22}(x,t) \end{pmatrix}, \quad B = \begin{pmatrix} b_{11}(x,t) & b_{12}(x,t) \\ b_{21}(x,t) & b_{22}(x,t) \end{pmatrix} \text{ and}$$

$$F = (F_1(x,t), F_2(x,t))^T.$$

We prescribe boundary conditions

$(3.3.1b)$ $M\, w(-1,t) = g(t)$

$(3.3.1c)$ $P\, w(1,t) = h(t)$

and initial condition

$(3.3.1d)$ $w(x,-1) = f(x)$.

Here M is a $\ell \times 2$ and P is a $\diamond \times 2$ matrix , where $0 \leq \ell ,\diamond \leq 2$. $\ell=0$ means there are no boundary conditions at x = -1 and $\diamond = 0$ that there are no boundary conditions at x = 1.

We use the central differenced discretization of

$$(3.3.2) \qquad w_t^{(n)} - A(x,t)w_x^n - B(x,t)w^n = \rho^{(n)}(x,t)$$

at interior points of the space-time square, to advance the solution from time level t_{j+1} to t_j.

As we have seen in the scalar case the equation uses information from a 6 point stencil. Thus to advance from time level t_{j+1} to t_j we have to solve a block tridiagonal system. To initialize the procedure we impose the initial condition

$$w_{i,2q}^{(n)} = \tau_i^{(n)} \qquad\qquad 0 \leq i \leq 2q.$$

It is evident that block-tridiagonal matrix solver constitutes the major portion of the numerical computation of the standard implicit algorithm. Equation (3.3.2) produces a 2 × 2 block structure for the implicit operator. T.H Pulliam and Chaussee (1981) have given an algorithm which transforms the coupled system of equations into an uncoupled diagonal form that requires considerably less computational work.

We describe this algorithm in brief for the system case. An implicit approximate factorization scheme for the system can be written as

$$(3.3.3) \qquad (I - A_{ij} \frac{\delta}{\delta x})\Delta w_j = R_{ij} .$$

Where $\Delta w_j = w_{j+1} - w_j$, and $A_{ij} = A(x_i,t_j)$,

as we can handle the lower order term explicitly. Here $\frac{\delta}{\delta x}$ denotes the centered difference approximation to the differential operator $\frac{\partial}{\partial x}$.

The matrix A_{ij} has a set of real eigenvalues and a complete set of eigenvectors , hence a similarity transformation can be used to diagonalize A_{ij}

(3.3.4) $A_{ij} = T_{ij} \Lambda_{ij} T_{ij}^{-1}$.

and so we write (3.3.3) as

(3.3.5) $(T_{ij}T_{ij}^{-1} - \Delta t \; T_{ij} \Lambda_{ij} T_{ij}^{-1} \frac{\delta}{\delta x}) \; \Delta w_j = R_{ij}$.

The modified form of the above equation is constructed by moving T outside   the difference operator $\frac{\delta}{\delta x}$. This results in the diagonal form of the algorithm

$$T_{ij}(\; I \; - \; \Delta t \; \Lambda_{ij} \frac{\delta}{\delta x}) \; T_{ij}^{-1} \Delta w_j = R_{ij} \; .$$

The modification has introduced an error, but T.H Pulliam and Chaussee have shown that the error introduced by the diagonalization is first order in time. The new implicit operator ($I - \Delta t \; \Lambda_{ij} \frac{\delta}{\delta x}$) is still block tridiagonal, but now blocks are diagonal in form so that the operator reduces to two independent scalar tridiagonal operators.

Numerical treatment of Boundary conditions

A correct treatment of the boundary conditionsis essential for an effective spectral calculation. If incorrect boundary conditions are imposed on the numerical scheme the resulting errors will propagate into thecomputational domain.   If these

errors propagateand/or grow sufficiently rapidly, they will destroy the solution.

Since the system is hyperbolic, A has real eigenvalues and a complete set of eigenvectors. So there exists a matrix T such that TAT^{-1} is diagonal. Equation (3.3.1a) can be rewritten as

$$T \, w_t - T \, A \, T^{-1} T \, w_x - T \, B \, T^{-1} T \, w = T \, F \quad , \text{or}$$

$$\tilde{w}_t - \Lambda \, \tilde{w}_x - \tilde{B} \, \tilde{w} = \tilde{F} \; .$$

Here

$$\tilde{w} = T \, w \qquad , \text{with}$$

$$\tilde{w} = (\, \tilde{w}_1 , \tilde{w}_2 \,) \; ,$$

$$\Lambda = T \, A \, T^{-1} \quad ,$$

$$\tilde{B} = T_t T^{-1} - \Lambda \, T_x T^{-1} - T \, B \, T^{-1} \quad , \text{and}$$

$$\tilde{F} = T \, F \; .$$

The variables $\tilde{w}_1$ and $\tilde{w}_2$ are called characteristic variables.

Assume $\tilde{w}_1$ is an inflow variable and $\tilde{w}_2$ is an outflow variable at $x = -1$. Then the boundary operator M would be of the form

$$(3.3.6) \qquad Mw(-1,t) = \tilde{w}_1(-1,t) - \alpha(t) \, \tilde{w}_2(-1,t)$$

where $\alpha(t)$ is a function of t. Hence the boundary condition at $x = -1$ could be written as

$$M \, w(-1,t) = g(t).$$

For the outflow variable we impose the partial differential

equation at the boundary implicitly. If we were to impose (3.3.6)
in the form

(3.3.7) $$\tilde{w}_1(-1,t_j) - \alpha(t_j)\,\tilde{w}_2(-1,t_j) = g(t_j)$$

the difference equation would no longer decouple into a set of
tridiagonal equations but instead would become block tridiagonal.
We can get around this problem by an approximate treatment of
the boundary condition. In (3.3.7) we approximate the unknown
value of $\tilde{w}_2(t_j)$ by using either

a) extrapolation ,or

b) an explicit finite difference discretization of the partial
differential equation.

It is easy to show that both these techniques, which we
describe below , are GKSO stable for a uniform mesh .

a(i) Zeroth order extrapolation

Here we simply put $\tilde{w}_2(-1,t_j) = \tilde{w}_2(-1,t_{j+1})$ for $2q-1 \geq j \geq 1$
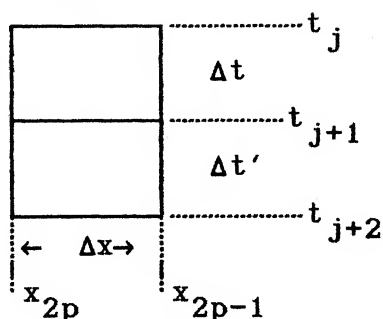
a(ii) First order extrapolation

fig 3.3.1

We define

$$\tilde{w}_2(-1,t_{2q-1}) = \tilde{w}_2(-1,t_{2q}) \quad \text{,and}$$

$$\tilde{w}_2(-1,t_j) = \tilde{w}_2(-1,t_{j+2}) + (\Delta t' + \Delta t) \times \frac{\tilde{w}_2(-1,t_{j+1}) - \tilde{w}_2(-1,t_{j+2})}{\Delta t}$$

$$0 \leq j \leq 2q-2.$$

b) Explicit difference scheme

Here we use an explicit difference scheme to compute $\tilde{w}_1(-1,t_j)$ from the values of $w(-1,t_{j+1})$ and $w(x_{2p-1},t_{j+1})$. Since the boundary condition (b) does not give good results we omit to describe it in detail.

Computational Results for the System case.

Example 3.3.1

$$A = \begin{pmatrix} -0.5+0.01\times\sin(\pi x) & 0.01 \\ 0.5 & 0.5+0.05\times\cos(\pi(x+t)) \end{pmatrix} , \quad B = 0$$

and $F(x,t) = \{ f_1(x,t), f_2(x,t) \}$

If our characteristic variables are $\tilde{u}_1(x,t)$ and $\tilde{u}_2(x,t)$ then $\tilde{u}_1(x,t)$ is an inflow variable at $x = 1$ and $\tilde{u}_2(x,t)$ is an inflow variable at $x = -1$.

Case I

Our solution is

$$u(x,t) = \begin{pmatrix} \sin(\frac{\pi}{33} x e^t) \\ \cos(2x - 3t) \times \epsilon \end{pmatrix}_{\epsilon = 0.1}$$

and the boundary conditions are of the form

$$\tilde{u}_2(-1,t) - 2 \times \sin(t) \tilde{u}_1(-1,t) = g(t) ,$$

$$\tilde{u}_1(1,t) - e^t \tilde{u}_2(1,t) = h(t)$$

with initial data

$$u(x,-1) = f(x) .$$

We omit to write the rather involved expressions for F,g,h,f .

Case II

We choose

$$u(x,t) = \begin{pmatrix} \sin(\frac{\pi}{16} \ x \ e^t) \\ \cos(2x - 3t) \times \in \end{pmatrix} \quad \in \ = \ 0.1$$

Then boundary conditions are

$$\tilde{u}_2(-1,t) - \tilde{u}_1(-1,t) = g(t) \ ,$$

$$\tilde{u}_1(1,t) - \tilde{u}_2(1,t) = h(t)$$

with initial data

$$u(x,-1) = f(x).$$

Number of collocation points is 33,

	A		B	
	Iterations	Error	Iterations	Error
Case I	4 56	9.248×10^{-3} 5.276×10^{-10}	2 50	2.300×10^{-3} 5.529×10^{-10}
Case II	5 57	8.868×10^{-3} 4.319×10^{-10}	2 51	2.890×10^{-3} 2.383×10^{-10}

Table 3.3.1

In A and B the value of the inflow variable at the boundary is obtained using boundary conditions a(i) and a(ii) respectively.

Example 3.3.2

$$A = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & -0.5 \end{pmatrix}$$

Case I $u_1(x,t) = x^3 + x \ t^2 + 100xt^7 + \cos(\ t \ )$

$u_2(x,t) = t^4 + (x + t)^3 + \sin(\ x \ )$

Case II $\quad u_1(x,t) = \cos\left(\frac{\pi}{100} \sin(x + t)\right)$

$\quad\quad\quad u_2(x,t) = t^4 + (x + t)^3 + \sin(x)$

Case III $\quad u_1(x,t) = \cos\left(\frac{\pi}{100} \sin(x + t)\right)$

$\quad\quad\quad u_2(x,t) = \sin(2x - t)$

Number of collocation points is 33,

	A		B	
	Iterations	Error	Iterations	Error
Case I	10	3.387×10^{-02}	2	7.348×10^{-03}
	25	2.048×10^{-07}	12	2.078×10^{-07}
Case II	45	7.521×10^{-03}	2	4.135×10^{-04}
	99	1.267×10^{-11}	50	9.393×10^{-13}
Case III	2	1.274×10^{-00}	2	8.812×10^{-05}
	25	6.163×10^{-01}	42	5.435×10^{-13}

Table 3.3.2

In A and B the value of the inflow variable at the boundary is obtained using boundary conditions (b) and a(ii) respectively. From table (3.2.2) we conclude boundary condition (b) gives poor results in general.

3.4. Preconditioning for Nonlinear Problems

We now consider the nonlinear IBVP

(3.4.1a) $\quad\quad u_t - A(u) u_x = F(x,t) \quad , \quad -1 \leq x \leq 1 , \quad -1 \leq t \leq 1,$

with boundary conditions

(3.4.1b) $\quad\quad\quad Mu(-1,t) = g(t) \quad , \quad -1 \leq t \leq 1 ,$

(3.4.1c) $\quad\quad\quad Pu(1,t) = h(t) \quad , \quad -1 \leq t \leq 1 ,$

and initial condition

(3.4.1d) $\quad\quad\quad u(x,-1) = f(x) \quad , \quad -1 \leq x \leq 1 ,$

and we assume that the solution u is smooth. We chose as our approximate solution $u^{p,q} \in (S^{p,q})^k$ which minimizes

$$H^{p,q}(v^{p,q}) = \frac{\pi^2}{4pq} \sum_{j=0}^{2q} \sum_{i=0}^{2p} ||(v_t^{p,q} - A(v^{p,q})v_x^{p,q} - F)(x_i^{2p}, t_j^{2q})||^2$$

$$+ \frac{\pi}{2q} \sum_{j=0}^{2q} ||(Mv^{p,q}(-1, t_j^{2q}) - g(t_j^{2q})||^2$$

(3.4.2)

$$+ \frac{\pi}{2q} \sum_{j=0}^{2q} ||(Pv^{p,q}(1, t_j^{2q}) - h(t_j^{2q})||^2$$

$$+ \frac{\pi}{2p} \sum_{i=0}^{2p} ||(v^{p,q}(x_i^{2p}, -1) - f(x_i^{2p})||^2$$

over all $v^{p,q} \in (S^{p,q})^k$ .Clearly, this gives rise to a nonlinear least – squares problems , which we may write as

(3.4.3) $L^{sp}(U)U = Z$.

We solve this nonlinear least – squares problem by the preconditioned residual minimization method as before. We outline the main steps below :

1)   To get an initial guess $U^{(0)}$ for the solution we let $V^{(0)}$ be the solution obtained on the finer mesh with $(2p+1) \times (2q+1)$ points by using a first or second order finite difference solver for the nonlinear IBVP (3.4.1) . Then we obtain $U^{(0)}$ from $V^{(0)}$ by truncating the highest half of its frequency components as before.

2)   Suppose at the nth stage of the iteration we have an approximate solution $U^{(n)}$ corresponding to $u^{(n)}(x,t)$, We can now calculate the residuals

$$\rho^{p,q}(x,t) = u_t^{(n)} - A(u^{(n)})u_x^{(n)} - \overline{F}^{2p,2q}(x,t) ,$$

$$\sigma^q(t) = M^q u^{(n)}(-1,t) - \overline{g}^{2q}(t) ,$$

(3.4.4) $\qquad \eta^q(t) = p^q u^{(n)}(1,t) - \overline{h}^{2q}(t) ,$

$$\tau^p(x) = u^{(n)}(x,-1) - \overline{f}^{2p}(x) .$$

We wish to find a correction $v(x,t)$ to $u^n(x,t)$, corresponding to the function $v(x,t) \in (S^{p,q})^k$ we have the vector V. Then $v(x,t)$ should approximately satisfy

(3.4.5a) $v_t - A(u^n)v_x - A(u^n)_x v = \rho^{(n)}(x,t)$,$-1 \le t \le 1$,
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad -1 \le x \le 1$,

(3.4.5b) $\left(\left[\dfrac{\partial M}{\partial u} \right]_{u=u^n} v \right)(-1,t) = \sigma^{(n)}(t)$,$-1 \le t \le 1$,

(3.4.5c) $\left(\left[\dfrac{\partial P}{\partial u} \right]_{u=u^n} v \right)(1,t) = \eta^{(n)}(t)$,$-1 \le t \le 1$,

(3.4.5d) $\qquad\qquad\qquad v(x,-1) = \tau^{(n)}(x)$,$-1 \le x \le 1$,

which is obtained by linearising (3.4.1) about $u^{(n)}$. Thus v can be obtained as the solution of a linear IBVP . Hence we can use the preconditioning techniques already described to obtain a correction V. Specifically, let W be the solution obtained on the finer mesh by a finite difference solver for (3.4.5), then V is obtained by truncating the highest half of the frequency components of W. We compute the relaxation factor ω_n so as to minimize the residual

$$H^{p,q}(\omega) = \frac{\pi^2}{4pq} \sum_{j=0}^{2q} \sum_{i=0}^{2p} ||(\omega\, v_t - \omega\, A(u^{(n)}) v_x - \rho^{(n)})(x_i^{2p}, t_j^{2q})||^2$$

$$+ \frac{\pi}{2q} \sum_{j=0}^{2q} ||\left(\omega \left[\frac{\partial M}{\partial u}\right]_{u=u^n} v\right)(-1, t_j^{2q}) - \sigma^{(n)}(t_j^{2q})||^2$$

(3.4.6)

$$+ \frac{\pi}{2q} \sum_{j=0}^{2q} ||\left(\omega \left[\frac{\partial P}{\partial u}\right]_{u=u^n} v\right)(1, t_j^{2q}) - \eta^{(n)}(t_j^{2q})||^2$$

$$+ \frac{\pi}{2p} \sum_{j=0}^{2p} ||(\omega\, v(x_i^{2p}, -1) - \tau^{(n)}(x_i^{2p})||^2 \ .$$

We then define

$$U^{(n+1)} = U^{(n)} + \omega_n V.$$

We remark that our numerical experiments indicate that it is enough to consider v as an approximate solution to the partial differential equation

$$v_t - A(u^{(n)})\, v_x = \rho^{(n)} \qquad , -1 \leq t \leq 1 \ , \ -1 \leq x \leq 1$$

along with the initial and boundary conditions (3.4.5b - 3.4.5d) to obtain convergence of the numerical scheme .

Computational Results for the Nonlinear case

Example 3.4.1

We consider the Burger's Equation

(3.4.7a) $\qquad u_t + u\, u_x = F(x,t)$,

with $\qquad u(x,t) = -10 + \sin(\frac{\pi}{16} x\, e^t).$

Then we have to impose the boundary condition

(3.4.7b) $u(-1,t) = g(t)$, $-1 \leq t \leq 1$,

and no boundary condition at $x = 1$.

The initial condition is

(3.4.7c) $u(x,-1) = f(x)$.

Example 3.4.2

We consider the isentropic Euler equation

$$(3.4.8) \qquad \begin{pmatrix} u \\ \rho \end{pmatrix}_t = \begin{pmatrix} u & \gamma\rho^{\gamma-2} \\ \gamma\rho & u \end{pmatrix} \begin{pmatrix} u \\ \rho \end{pmatrix}_x$$

with

$u(x,t) = x^2 + 10 + \cos(\frac{\pi}{16} \sin(x+t))$

$\rho(x,t) = x^2 + 2 + \sin(2x-t)^2$ and $\gamma = 2.0$

The flow is supersonic and so we have to specify no boundary condition at $x = -1$ and both ρ and u at $x = 1$.

Number of collocation points is 33.

Example No.		
	Iterations	Error
3.4.1	6	6.487×10^{-03}
	25	1.682×10^{-10}
3.4.2	6	6.463×10^{-3}
	25	1.571×10^{-10}

GALERKIN – COLLOCATION METHOD
FOR LEGENDRE POLYNOMIALS

4.1 Introduction

Galerkin – Collocation method can be applied to general Geggenbauer polynomials. In chapter II and chapter III we have applied this method to Chebyshev polynomials. This chapter deal with Legendre polynomials and the results obtained are similar to the one obtained in chapter II. We now outline the contents of this chapter. In section two we discuss the basic properties of legendre polynomials and in the third section we have applied the Galerkin – Collocation method to Legendre polynomials. The computational results for scalar problemsare presented in section four.

4.2 Properties of Legendre polynomial

We present here a collection of essential formulae for Legendre polynomials which are needed in the later part of this chapter. For proofs, the reader may refer to Szegö (1939). Legendre polynomials { $L_k(x)$, k=0,1,...,} are the eigen functions of the Strum Liouville problem

$$-(P(x)U'(x))' + Q(x)U(x) = \lambda\ \omega(x)U(x)$$

in the interval (-1,1) with $P(x) = 1 - x^2$, $Q(x) = 0$ and $\omega(x) = 1$. $L_k(x)$ is an even function of x if k is even and is an odd function of x if k is odd.

If $L_k(x)$ is normalized so that $L_k(1) = 1$ then for any k

$$L_k(x) = \frac{1}{2^k} \sum_{\ell=0}^{[k/2]} (-1)^\ell \, {}^kC_\ell \, {}^{2k-2\ell}C_k \, x^{k-2\ell} \qquad ,$$

where [k/2] denote the integral part of k/2.

Legendre polynomials are orthogonal with respect to the weight function

$$w(x) = 1. \qquad \int_{-1}^{1} (L_k(x))^2 \, dx = (k + \frac{1}{2})^{-1}.$$

The expansion of $u(x) \in L^2(-1,1)$ in term of L_k's is

$$u(x) = \sum_{k=0}^{\infty} \hat{u}_k \, L_k(x)$$

where $\hat{u}_k$ given by

$$\hat{u}_k = (k+1/2) \int_{-1}^{1} u(x) \, L_k(x) \, dx \quad ,$$

are called Legendre coefficients.

Legendre-Gauss-Lobbato points defined as $x_0^N = -1$, $x_N^N = 1$ and $x_j = (1,2,\ldots,N-1)$ are the roots of polynomial $L_n'(x)$. We now consider the discrete Legendre transforms

$$\bar{u}(x) = \sum_{k=0}^{N} \bar{u}_k \, L_k(x) \quad .$$

$\bar{u}(x)$ is the unique polynomial of degree $\leq N$ such that $\bar{u}(x_j^N) = u(x_j^N)$ $\forall$ $x_j^N = 0,1,\ldots,N$ where x_j^N are Legendre-Gauss-Lobbato points. $\bar{u}_k$ are called discrete polynomial coefficients of u(x). The inverse relationship is given by

$$\bar{u}_k = \frac{1}{\gamma_k} \sum_{j=0}^{N} u(x_j^N) \, L_k(x_j^N) \, w_j^N$$

where w_j^N and γ_k are quadrature weights and normalizing factor, respectively and are given by the formulas

$$w_j^N = \frac{1}{N(N+1) \, (L_N(x_j))^2} \quad \forall \, j = 0,1,\ldots,N$$

and
$$\gamma_k = \sum_{j=0}^{N} (L_k(x_j^N))^2 \, w_j^N \qquad \forall \; k = 0, 1, \ldots, N \; .$$

4.3 Discussion of Method and Theoretical Results

We shift our initial time t = 0, to t = -1.

Consider a well posed IBVP

(4.3.1a) $\qquad Lu(x,t) = F(x,t) \qquad -1 \le x \le 1, \; -1 \le t \le 1$

with boundary conditions

(4.3.1b) $\qquad Mu(-1,t) = g(t) \qquad -1 \le t \le 1$

(4.3.1c) $\qquad Pu(1,t) = h(t) \qquad -1 \le t \le 1$

and initial condition

(4.3.1d) $\qquad u(x,-1) = f(x) \qquad -1 \le x \le 1$

which is structurally stable. The operator L, M, P and function F(x,t), g(t), h(t) and f(x) have the same meaning as described in section 2.2.

Let $S^{p,q}$ be the set of polynomial $w^{p,q}(x,t)$ of the form

(4.3.2) $\qquad w^{p,q}(x,t) = \sum_{j=0}^{q} \sum_{i=0}^{p} a_{ij} \, L_i(x) \, L_j(t)$

with scalar coefficients a_{ij} and L_i be the ith Legendre polynomial. Similarly we shall define by $(S^{p,q})^k$ the set of polynomials $w^{p,q}$ of the form (4.3.2) if the coefficients a_{ij} are vectors of length k. Henceforth we shall assume that there exist a constant λ such that

$$\frac{1}{\lambda} \le \frac{p}{q} \le \lambda$$

Now we define the interpolation operator $I^{p,q}$ which takes a continuous function r(x,t) defined on [-1,1]×[-1,1] and projects it into $S^{p,q}$. Thus

(4.3.3) $\qquad I^{p,q} r(x,t) = \sum_{j=0}^{q} \sum_{i=0}^{p} a_{ij} \, L_i(x) \, L_j(t) = \bar{r}^{p,q}(x,t)$

is the unique polynomial belonging to $S^{p,q}$ which interpolates $r(x,t)$ at the $(p+1)\times(q+1)$ point $\{(x_i^p, t_j^q)\}_{\substack{i=0,1,\ldots,p \\ j=0,1,\ldots,q}}$. Here the point x_i^p , t_j^q are Legendre-Gauss-Lobbato points.

In much the same way we can define interpolation operator I^l for one space dimension which takes a continuous function $s(y)$ defined in $[-1,1]$ and project it into the space of polynomial of degree $\leq l$. Thus

(4.3.4)
$$I^l s(y) = \sum_{i=0}^{l} b_i L_i(y) = \bar{s}^l(y)$$

is the unique polynomial of degree $\leq l$ which interpolates $s(y)$ at the $(l+1)$ Legendre-Gauss-Lobbato points $\{y_i\}_{i=0,1,\ldots,l}$.

Using these interpolation operators we define a filtered version of the differential operator

(4.3.5a)
$$L^{p,q} \tilde{u}(x,t) = \tilde{u}_t(x,t) - \bar{A}^{p,q} \tilde{u}_x(x,t) - \bar{B}^{p,q} \tilde{u}(x,t) = F(x,t)$$

with boundary conditions

(4.3.5b)
$$M^q \tilde{u}(-1,t) = g(t)$$

(4.3.5c)
$$P^q \tilde{u}(1,t) = h(t)$$

and initial condition

(4.3.5d)
$$\tilde{u}(x,-1) = f(x) .$$

The above IBVP is well posed if we choose p,q large enough. Since (4.3.5) can be regarded as a perturbation of (4.3.1) the following energy estimate

$$\int_{-1}^{1} \int_{-1}^{1} \| u^{p,q}(x,t) \|^2 dx\, dt + \int_{-1}^{1} \| u^{p,q}(-1,t) \|^2 dt$$

$$+ \int_{-1}^{1} \| u^{p,q}(1,t) \|^2 dt + \int_{-1}^{1} \| u^{p,q}(x,1) \|^2 dx$$

$$\leq C \left[\int_{-1}^{1} \int_{-1}^{1} \| L^{p,q} u^{p,q}(x,t) \|^2 dx\, dt + \int_{-1}^{1} \| u^{p,q}(x,-1) \|^2 dx \right.$$

$$(4.3.6) \qquad + \int_{-1}^{1} \| M^q u^{p,q}(-1,t) \|^2 dt + \int_{-1}^{1} \| P^q u^{p,q}(1,t) \|^2 dt \Big],$$

holds for p and q large enough , with some constant C. Henceforth we shall let C denote a generic constant .

From the above the inequality

$$\int_{-1}^{1} \int_{-1}^{1} \|u^{p,q}(x,t)\|^2 dx dt \leq C \Big[\int_{-1}^{1} \int_{-1}^{1} \|L^{p,q} u^{p,q}(x,t)\|^2 dx dt$$

$$+ \int_{-1}^{1} \|u^{p,q}(x,-1)\|^2 dx + \int_{-1}^{1} \|M^q u^{p,q}(-1,t)\|^2 dt$$

$$(4.3.7) \qquad + \int_{-1}^{1} \|P^q u^{p,q}(1,t)\|^2 dt \Big],$$

follows.

Notice that if $u^{p,q}(x,t) \in (S^{p,q})^k$ then

$$L^{p,q} u^{p,q}(x,t) \in (S^{2p,2q})^k \ ,$$

$$M^q u^{p,q}(-1,t) \in (S^{2q})^{\ell} \ ,$$

$$P^q u^{p,q}(1,t) \in (S^{2q})^s \ , \quad \text{and}$$

$$u^{p,q}(x,-1) \in (S^p)^k \ ,$$

and this suggests that we should accordingly filter our data. Let

$$\overline{F}^{2p,2q}(x,t) = I^{2p,2q} F(x,t) \ ,$$

$$\overline{g}^{2q}(x,t) = I^{2q} g(t) \ ,$$

$$\overline{h}^{2q}(t) = I^{2q} h(t) \ , \quad \text{and}$$

$$\overline{f}^{2p}(x) = I^{2p} f \ ,$$

be filtered representations of the data. If we substitute our approximate solution into the IBVP the residuals

$$\rho^{p,q}(x,t) = L^{p,q} u^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t) \ ,$$

$$\sigma^q(t) \quad = \quad M^q \, u^{p,q}(-1,t) - \overline{g}^{2q}(t) \quad ,$$

$$(4.3.8) \qquad \eta^q(t) \quad = \quad P^q \, u^{p,q}(1,t) - \overline{h}^{2q}(t) \quad ,$$

$$\tau^p(x) \quad = \quad u^{p,q}(x,-1) - \overline{f}^{2p}(x) \quad ,$$

in general not be zero. We would like to choose an approximate solution $u^{p,q}(x,t)$ so that it makes these residuals as small as possible and for this we define a functional which will measure the size of the residuals.

Accordingly we define a functional

$$H^{p,q}(v^{p,q}) = \int_{-1}^{1} \int_{-1}^{1} ||L^{p,q} \, v^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t)||^2 \, dx \, dt$$

$$+ \int_{-1}^{1} ||M^q v^{p,q}(-1,t) - \overline{g}^{2q}(t)||^2 dt \, + \int_{-1}^{1} ||P^q v^{p,q}(1,t) - \overline{h}^{2q}(t)||^2 dt$$

$$(4.3.9) \qquad\qquad\qquad\qquad + \int_{-1}^{1} || \, v^{p,q}(x,-1) - \overline{f}^{2p}(x) \, ||^2 dx,$$

where $v^{p,q}(x,t) \in (S^{p,q})^k$.

We choose as our approximate solution the unique $u^{p,q} \in (S^{p,q})^k$ which minimizes a functional $H^{p,q}(v^{p,q})$ over all $v^{p,q}$, where $H^{p,q}(v^{p,q})$ is essentially equivalent to $H^{p,q}(v^{p,q})$.

Now we observe that

$$\rho^{p,q}(x,t) \quad = \quad L^{p,q} \, v^{p,q}(x,t) - \overline{F}^{2p,2q}(x,t) \in (\, S^{2p,2q})^k,$$

$$\sigma^q(t) \quad = \quad M^q \, v^{p,q}(-1,t) - \overline{g}^{2q}(t) \in (S^{2q})^\ell,$$

$$\eta^q(t) \quad = \quad P^q \, v^{p,q}(1,t) - \overline{h}^{2q}(t) \in (S^{2q})^s , \quad \text{and}$$

$$\tau^p(x) \quad = \quad v^{p,q}(x,-1) - \overline{f}^{2p}(x) \in (S^{2p})^k ,$$

and so we can exactly evaluate the integrals in (4.3.9) by using the very highly accurate Gauss quadrature rules. In particular, for the Legendre-Gauss-Lobatto rule we have that if s(y) is a

polynomial of degree $\leq 2N - 1$ then

(4.3.10)
$$\int_{-1}^{1} s(y)\, dy = \sum_{j=0}^{N} w_j^N\, s(y_j^N)$$

where the points y_j^N are Legendre-Gauss-Lobbato points and the weights w_j^N are corresponding weights. Moreover, the inequality

(4.3.11)
$$\frac{1}{3} \sum_{j=0}^{N} w_j^N\, |s(y_j^N)|^2 \leq \int_{1}^{1} s^2(y)\, dy \leq \sum_{j=0}^{N} w_j^N\, |s(y_j^N)|^2$$

holds if $s(y)$ is a polynomial of degree $\leq N$.

We can therefore replace the functional $\bar{H}^{p,q}(v^{p,q})$ we are trying to minimize by an equivalent functional

$$
\begin{aligned}
H^{p,q}(v^{p,q}) = \sum_{j=0}^{2q} \sum_{i=0}^{2p} &\|L^{p,q}\, v^{p,q}(x_i^{2p}, t_j^{2q}) - \bar{F}^{2p,2q}(x_i^{2p}, t_j^{2q})\|^2\, w_i^{2p}\, w_j^{2q} \\
&+ \sum_{j=0}^{2q} \| M^q\, v^{p,q}(-1, t_j^{2q}) - \bar{g}^{2q}(t_j^{2q}) \|^2\, w_j^{2q} \\
&+ \sum_{j=0}^{2q} \| P^q\, v^{p,q}(1, t_j^{2q}) - \bar{h}^{2q}(t_j^{2q}) \|^2\, w_j^{2q} \\
&+ \sum_{i=0}^{2p} \| v^{p,q}(x_i^{2p}, -1) - \bar{f}^{2p}(x_i^{2p}) \|^2\, w_i^{2p}
\end{aligned}
$$

(4.3.12)

In fact, using (4.3.11) we conclude that

$$\bar{H}^{p,q}(v^{p,q}) \leq H^{p,q}(v^{p,q}) \leq 9\, \bar{H}^{p,q}(v^{p,q}).$$

We choose an approximate solution $u^{p,q} \in (S^{p,q})^k$ which minimizes $H^{p,q}$.

In other words, our solution $u^{p,q}$ is given by a least-squares solution to the overdetermined system of equations

$$\left(w_i^{2p}\, w_j^{2q} \right)^{1/2} \left\{ L^{p,q}\, u^{p,q} - F)(x_i^{2p}, t_j^{2q}) \right\} = 0,$$

$$0 \leq i \leq 2p\ ,\ 0 \leq j \leq 2q,$$

$$\left(w_j^{2q}\right)^{1/2} \left\{ M^q(t_j^{2q})\ u^{p,q}(-1,t_j^{2q}) - g(t_j^{2q})\right\}= 0,$$

(4.3.13)
$$0 \le j \le 2q,$$

$$\left(w_j^{2q}\right)^{1/2} \left\{ P^q(t_j^{2q})\ u^{p,q}(-1,t_j^{2q}) - h(t_j^{2q})\right\} = 0,$$

$$0 \le j \le 2q,$$

$$\left(w_i^{2p}\right)^{1/2} \left\{ u^{p,q}(x_i^{2p},-1) - f)(x_i^{2p})\right\} = 0,$$

$$0 \le i \le 2p.$$

Here, we have used the fact that $\overline{F}^{2p,2q}(x_i^{2p},t_j^{2q}) = F(x_i^{2p},t_j^{2q})$ etc. and so can work with point values of the original data. We may write the system (4.3.13) in the form

(4.3.14)
$$D^{p,q}\ U^{p,q} = Z^{p,q},$$

In chapter II we have shown the filtering can be dispensed with and it suffices to collocate the partial differential equation and initial and boundary conditions at the over determined set of points. This result hold even in the case of Legendre polynomials. Thus the system (4.3.14) may be written as

(4.3.15)
$$\widetilde{D}^{p,q}\ \widetilde{U}^{p,q} = \widetilde{Z}^{p,q},$$

where $\widetilde{U}^{p,q}$ is the least – square solution to the unfiltered system of equations

$$\left(w_i^{2p}\ w_j^{2q}\right)^{1/2} \left\{ L\ \widetilde{u}^{p,q}(x_i^{2p},t_j^{2q}) - F(x_i^{2p},t_j^{2q})\right\} = 0,$$

$$0 \le i \le 2p\ ,\ 0 \le j \le 2q,$$

$$\left(w_j^{2q}\right)^{1/2} \left\{ M(t_j^{2q})\ \widetilde{u}^{p,q}(-1,t_j^{2q}) - g(t_j^{2q})\right\} = 0,$$

(4.3.16)
$$0 \le j \le 2q,$$

$$\left(w_j^{2q}\right)^{1/2} \left\{ P(t_j^{2q})\ \widetilde{u}^{p,q}(-1,t_j^{2q}) - h(t_j^{2q})\right\} = 0,$$

$$0 \le j \le 2q,$$

$$\left(w_i^{2p}\right)^{1/2} \left\{\tilde{u}^{p,q}(x_i^{2p},-1) - f)(x_i^{2p})\right\} = 0,$$

$$0 \le i \le 2p.$$

It is simple to prove the analog of the lemmas 2.2.1 and 2.2.2 and of theorem 2.2.1 for the Legendre case.

4.4 Numerical Results

The following computational results were obtained using the preconditioned residual minimization method discussed in chapter III.

Computation for scalar case

Example 4.4.1

$$u_t - a(x,t)\,u_x - b(x,t)\,u = F(x,t)$$

$$-1 \le x \le 1,\ -1 \le t \le 1$$

where

$$a(x,t) = x + \frac{1}{2}\sin(\pi(x+t)) \ ,\ \ b(x,t) = \sin\frac{(x+t)}{2}\ ,$$

$$F(x,t) = \left\{\frac{1}{2}C_n e^t \sin(\pi(x+t))\,\cos(C_n x e^t) + \sin\frac{(x+t)}{2}\,\sin(C_n x e^t)\right\}\ ,$$

with initial conditions

$$u(x,-1) = \sin(C_n x e^{-1})\ ,$$

and boundary conditions

$$u(-1,t) = -\sin(C_n e^t)\ ,$$

$$u(1,t) = \sin(C_n e^t)\ .$$

The results obtained for three different values of C_n are shown in table 4.1.1.

N – Number of collection points

N	$C_n = \dfrac{\pi}{16}$	$C_n = \dfrac{\pi}{32}$	$C_n = \dfrac{\pi}{64}$
17	$\langle 2 \rangle$ (1.05×10^{-4}) $\langle 48 \rangle$ (6.87×10^{-9})	$\langle 1 \rangle$ (5.26×10^{-4}) $\langle 45 \rangle$ (9.27×10^{-9})	$\langle 1 \rangle$ (2.73×10^{-4}) $\langle 44 \rangle$ (6.11×10^{-9})

Table 4.1.1

where the number in the frist brackets denotes the interation
number at which the error given in the second brackets is
obtained.

CHAPTER V

SPECTRAL METHODS FOR PERIODIC INITIAL VALUE PROBLEMS WITH NONSMOOTH DATA

5.1 Introduction

In this chapter we consider hyperbolic initial value problems subject to periodic boundary conditions with nonsmooth data. We show that if we filter the data and solve the problem by the Galerkin-Collocation method , discussed in chapter II, then we can recover pointwise values with spectral accuracy , provided that the actual solution is piecewise smooth. For this we have to perform a local smoothing of the computed solution.

We now outline the contents of this chapter. In Section 2 we define the Sobolev spaces we shall work in and describe the energy estimates in negative Sobolev norms which are needed in this chapter. In Section 3 we briefly describe the Galerkin-Collocation method and prove that the error between the approximate solution computed by this method and the actual solution in a negative Sobolev norm decays at a rate which depends only on the order of the norm. In Section 4 we explain the filtering procedure proposed by Abarbanel, Gottlieb and Tadmor and show how it can be applied to the approximate solution we obtain by the Galerkin-Collocation method to recover pointwise values of the solution with spectral accuracy. Finally in Section 5 we present computational results for the proposed method.

5.2 Energy Estimates for Hyperbolic Initial Value Problems with Periodic Boundaries

We consider hyperbolic initial value problems with periodic boundary conditions. Here after x denotes the vector

$$x = (x_1, x_2, \ldots, x_d).$$

Let $\Omega = (0, 2\pi)^d$ be the space domain and $J = (-1, 1)$ be the time interval we are considering. Consider the IVP

$$(5.2.1) \qquad Lu = u_t - \sum_{i=1}^{d} A_i u_{x_i} - B u = F \quad \text{for } (x,t) \in \Omega \times J ,$$

$$u = f \qquad\qquad\qquad \text{for } (x,t) \in \Omega \times \{-1\}.$$

Here u is a vector valued function with values in R^P and A_i, B are matrix valued functions. Moreover A_i, B are smooth functions of x and t and periodic in x_j with period 2π, for all $j = 1, \ldots, d$, and f and F are periodic in each space coordinate with the same period but are not necessarily smooth.

Before we proceed to describe our numerical method and prove its convergence we need to review some a priori energy estimates which have been proved for the solutions of the system (5.2.1). The interested reader is referred to Rauch(1972) and Taylor(1981) for details.

Let u and v be vector valued functions of x and t and 2π-periodic in each space direction. Then we denote

$$(u,v)_{\Omega \times J} = \int \int_{\Omega \times J} u^* v \; dxdt , \text{ and}$$

$$\| u \|_{0, \Omega \times J} = \left(\int \int_{\Omega \times J} |u|^2 \; dxdt \right)^{1/2} .$$

Here |u| denotes the euclidean norm of u if u is a vector and |A| denotes the induced matrix norm if A is a matrix. Similarily we

denote

$$\| u \|_{s,\Omega \times J} = \left(\int\int_{\Omega \times J} \sum_{|\alpha|+\beta \leq s} |D_x^\alpha D_t^\beta u|^2 \, dxdt \right)^{1/2} \, ,$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ is a multi-index and

$$D_x^\alpha u = D_{x_1}^{\alpha_1} \ldots D_{x_d}^{\alpha_d} u.$$

In the same way we define

$$(u,v)_{\Omega \times \{\pm 1\}} = \int_{\Omega \times \{\pm 1\}} u^* v \, dx \, , \text{ and}$$

$$\| u \|_{0 \times \Omega \times \{\pm 1\}} = \left(\int_{\Omega \times \{\pm 1\}} |u|^2 \, dx \right)^{1/2}.$$

Let

$$\| u \|_{s,\Omega \times \{\pm 1\}} = \left(\int_{\Omega \times \{\pm 1\}} \sum_{|\alpha| \leq s} |D_x^\alpha u|^2 \, dx \right)^{1/2} \, ,$$

where α is a multi-index as above.

We can now state the a priori energy estimates conveniently in terms of the Sobolev norms we have just defined. Let ψ be the solution of the hyperbolic IVP with periodic boundary conditions

(5.2.2a) $L\psi = \phi$ for $(x,t) \in \Omega \times J$,

(5.2.2b) $\psi = \theta$ for $(x,t) \in \Omega \times \{-1\}$,

where ϕ and θ are smooth functions and periodic in space. Then for all integers $s \geq 0$ there exists a constant C_s , which depends only on the smoothness properties of A_i , B such that the estimate

(5.2.3) $\| \psi \|_{s,\Omega \times J} + \| \psi \|_{s,\Omega \times \{1\}}$

$$\leq C_s \left(\| \phi \|_{s,\Omega \times J} + \| \theta \|_{s,\Omega \times \{-1\}} \right)$$

holds. Henceforth we shall use C and C_s as generic constants. Next, we need to state a version of (5.2.3) for negative Sobolev

norms.

Let w be a function of x and t which is periodic in space. Let $H = \{ \phi : \phi$ is a smooth function of x and t which is periodic in x and has compact support in t $\}$. We define

$$\| w \|_{-s,\Omega\times J} = \sup_{\phi \in H} \frac{| (w,\phi)_{\Omega\times J} |}{\| \phi \|_{s,\Omega\times J}} ,$$

Then $H_{-s,\Omega\times J}$ is defined to be the completion of H with respect to the above norm. Similarily we define

$$\| w \|_{-s,\Omega\times\{-1\}} = \sup_{\phi \in H} \frac{| (w,\phi)_{\Omega\times\{-1\}} |}{\| \phi \|_{s,\Omega\times\{-1\}}} .$$

With these definitions we can now state the energy estimates in "negative" Sobolev norms. For any $s \geq 0$ there exists a constant C_s, which depends only on the smoothness properties of A_i , B such that

(5.2.4)
$$\| \psi \|_{-s,\Omega \times J} + \| \psi \|_{-s,\Omega\times\{1\}} \leq C_s (\| \phi \|_{-s,\Omega\times J} + \| \theta \|_{-s,\Omega\times\{-1\}}) ,$$

where ψ is the solution of (5.2.2), for all ϕ and θ. For the sake of completeness we shall provide the proof of (5.2.4) below , which is very similar to an analogous result proved by Rauch(1972).

We consider the following hyperbolic IVP with periodic boundary conditions

(5.2.5a) $\quad L^* w = -w_t + \sum_{i=1}^{d} \left(A_i^T w\right)_{x_i} - B^T w = \chi \quad$ for $(x,t) \in \Omega\times J$,

(5.2.5b) $\quad w = \mu \qquad\qquad\qquad\qquad\qquad$ for $(x,t) \in \Omega\times\{1\}$,

which is the adjoint of (5.2.2). Notice that for this problem we

let time run backwards. The following energy estimate is then valid for the solution w of the adjoint problem .

For every $s \geq 0$ there exists a constant C_s which depends only on the smoothness properties of A_i , B such that

$$(5.2.6) \qquad \| w \|_{s,\Omega \times J} + \| w \|_{s,\Omega \times \{-1\}}$$
$$\leq C_s (\| \chi \|_{s,\Omega \times J} + \| \mu \|_{s,\Omega \times \{1\}})$$

holds.

Let ψ be the solution of (5.2.2). An integration by parts yields

$$(5.2.7) \qquad (\psi, L^* w)_{\Omega \times J} = (L\psi, w)_{\Omega \times J} + (\psi, w)_{\Omega \times \{-1\}} - (\psi, w)_{\Omega \times \{1\}} ,$$

since the integrands are periodic in space.

Let w be the solution of the adjoint IVP with periodic boundary conditions

$$(5.2.8a) \qquad L^* w = \chi \qquad \text{for } (x,t) \in \Omega \times J ,$$

$$(5.2.8b) \qquad w = 0 \qquad \text{for } (x,t) \in \Omega \times \{1\} .$$

Then by (5.2.7) we have

$$(5.2.9) \qquad |(\psi, \chi)_{\Omega \times J}| \leq \| L\psi \|_{-s,\Omega \times J} \times \| w \|_{s,\Omega \times J}$$
$$+ \| \psi \|_{-s,\Omega \times \{-1\}} \times \| w \|_{s,\Omega \times \{-1\}} .$$

But using the estimate (5.2.6) we have

$$\| w \|_{s,\Omega \times J} + \| w \|_{s,\Omega \times \{-1\}} \leq C_s \| \chi \|_{s,\Omega \times J} ,$$

and this together with (5.2.9) gives

$$(5.2.10) \qquad |(\psi, \chi)_{\Omega \times J}| \leq C_s (\| L\psi \|_{-s,\Omega \times J}$$
$$+ \| \psi \|_{-s,\Omega \times \{-1\}}) \times \| \chi \|_{s,\Omega \times J} .$$

Thus from (5.2.10) we obtain

$$(5.2.11) \qquad \| \psi \|_{-s,\Omega \times J} \leq C_s (\| L\psi \|_{-s,\Omega \times J} + \| \psi \|_{-s,\Omega \times \{-1\}}) .$$

Next , let w˙ be the solution of the adjoint IVP with periodic boundary conditions

(5.2.12a) $L^* w = 0$ for $(x,t) \in \Omega \times J$,

(5.2.12b) $w = \mu$ for $(x,t) \in \Omega \times \{1\}$.

Then (5.2.7) takes the form

(5.2.13) $(\psi, \mu)_{\Omega \times \{1\}} = (L\psi, w)_{\Omega \times J} + (\psi, w)_{\Omega \times \{-1\}}$,

and by (5.2.6) the estimate

(5.2.14) $\| w \|_{s, \Omega \times J} + \| w \|_{s, \Omega \times \{-1\}} \leq C_s \| \mu \|_{s, \Omega \times \{1\}}$

is valid.

Now (5.2.13) and (5.2.14) give

$$| (\psi, \mu)_{\Omega \times \{1\}} | \leq C_s \; (\; \| L\psi \|_{-s, \Omega \times J}$$

$$+ \| \psi \|_{-s, \Omega \times \{-1\}} \;) \times \| \mu \|_{s, \Omega \times \{1\}} \; ,$$

from which we obtain

(5.2.15) $\| \psi \|_{-s, \Omega \times \{1\}} \leq C_s \; (\; \| L\psi \|_{-s, \Omega \times J} + \| \psi \|_{-s, \Omega \times \{-1\}} \;).$

Combining (5.2.11) and (5.2.15) we get (5.2.4).

5.3   Error Estimates for Blended   Fourier-Legendre   Methods
for Periodic Problems with Nonsmooth Data

Henceforth we shall take $\Omega = (0, 2\pi)$ since the results we state carry over to the general case $\Omega = (0, 2\pi)^d$ in a straightforward manner. We now introduce some notation. For each integer N we denote by Π^N the space of algebraic polynomials in the variable t of degree upto N. For each integer M we denote by S^M the space

$$S^M = \text{span} \{ e^{ikx} \mid -M \leq k \leq M \}.$$

Then we define the space $V^{M,N}$ as the tensor product

$$V^{M,N} = \{ \phi : \phi(x,t) = \sum_{n=0}^{N} \sum_{m=-M}^{M} a_{mn} e^{imx} L_n(t) \} ,$$

where $L_n(t)$ is the Legendre polynomial of degree n. Henceforth we shall assume that there exists a constant λ such that

$$1/\lambda \leq M/N \leq \lambda .$$

For any function w periodic in x , which also belongs to $L^2(\Omega \times J)$, let $P^{M,N}$ w denote the projection of w into $(V^{M,N})^P$, i.e.

$$P^{M,N}w = \sum_{n=0}^{N} \sum_{m=-M}^{M} w_{mn} e^{imx} L_n(t) ,$$

where

$$w = \sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} w_{mn} e^{imx} L_n(t) .$$

Henceforth we shall denote $P^{M,N}$ w by $\bar{w}^{M,N}$.

The following results are well known , (Canuto *et.al.*1988).

If $w \in H_{k,\Omega \times J}$ then

$$(5.3.1) \qquad \| w - \bar{w}^{M,N} \|_{0,\Omega \times J} \leq C N^{-k} \| w \|_{k,\Omega \times J} .$$

Moreover

$$(5.3.2) \qquad \| \bar{w}^{M,N} \|_{0,\Omega \times J} \leq \| w \|_{0,\Omega \times J} .$$

Also we have

$$(5.3.3) \qquad \| w - \bar{w}^{M,N} \|_{1,\Omega \times J} \leq C N^{2l-k} \| w \|_{k,\Omega \times J} ,$$

for all $0 \leq l \leq k$.

Next , we introduce the norm

$$\| w \|_{s,\infty,\Omega \times J} = \max_{\alpha+\beta \leq s} (\operatorname{ess\,sup}_{(x,t) \in \Omega \times J} | D_x^{\alpha} D_t^{\beta} w |).$$

Then we have

$$(5.3.4) \qquad \| w - \bar{w}^{M,N} \|_{1,\infty,\Omega \times J} \leq C N^{2l-k} \| w \|_{k,\infty,\Omega \times J} ,$$

for all $0 < l \leq k$.

If $s(x)$ is a periodic function belonging to $L^2(\Omega)$ we define

$$P^{M,0}s = \sum_{m=-M}^{M} s_m e^{imx} = \bar{s}^M,$$

where $s(x) = \sum_{m=-\infty}^{\infty} s_m e^{imx}$.

Similarily , if $h(t) \in L^2(J)$ we define

$$P^{0,N}h = \sum_{n=0}^{N} h_n L_n(t) = \bar{h}^N,$$

where $h(t) = \sum_{n=0}^{\infty} h_n L_n(t)$.

We have results similar to (5.3.1)-(5.3.3) for the above.

Let

$$\bar{A}^{M-1,N-1} = P^{M-1,N-1} A ,$$

$$\bar{B}^{M-1,N-1} = P^{M-1,N-1} B ,$$

$$\bar{F}^{2M-1,2N-1} = P^{2M-1,2N-1} F ,$$

$$\bar{f}^{2M-1} = P^{2M-1,0} f .$$

We define the differential operator

$$L^{M,N}w = w_t - \bar{A}^{M-1,N-1} w_x - \bar{B}^{M-1,N-1} w.$$

We choose as our approximate solution

$$v^{M,N} \in (V^{M,N})^p = \{ \phi : \phi(x,t) = \sum_{n=0}^{N} \sum_{m=-M}^{M} a_{mn} e^{imx} L_n(t) ,$$

$$a_{mm} \in R^p \}$$

which minimizes

(5.3.5) $$\mathscr{X}^{M,N}(w^{M,N}) = \int\int_{\Omega \times J} |L^{M,N} w^{M,N} - \bar{F}^{2M-1,2N-1}|^2 \, dxdt$$

$$+ \int_{\Omega \times \{-1\}} |w^{M,N}(x,-1) - \bar{f}^{2M-1}(x)|^2 \, dx$$

over all $w^{M,N} \in (V^{M,N})^p$.

The above problem reduces to obtaining a least-squares solution to an overdetermined set of equations obtained by collocating the

modified equation

$$L^{M,N} \ w^{M,N} = \bar{F}^{2M-1,2N-1},$$

and the initial conditions at an overdetermined set of points. We briefly explain this.

Let $\chi_i^M = \pi i / M$, $0 \leq i \leq 2M-1$,and let $\{\tau_j^N\}_{j=0,\dots,N}$ be the Gauss-Lobatto-Legendre points with $\tau_0 = -1$ and $\tau_N = 1$. Notice that

$$(L^{M,N} \ w^{M,N} - \bar{F}^{2M-1,2N-1}) \in (V^{2M-1,2N-1})^p ,$$

and

$$(w^{M,N}(x,-1) - \bar{f}^{2M-1}) \in (S^{2M-1})^p.$$

Hence we have that

$$\mathscr{X}^{M,N}(w^{M,N}) = \sum_{j=0}^{2N} \sum_{i=0}^{4M-1} \alpha_{ij}^{M,N} |(L^{M,N} \ w^{M,N} - \bar{F}^{2M-1,2N-1})(\chi_i^{2M},\tau_j^{2N})|^2$$

(5.3.6)
$$+ \sum_{i=0}^{4M-1} \beta_i^M \ | \ w^{M,N}(\chi_i^{2M},-1) - \bar{f}^{2M-1}(\chi_i^{2M})|^2 ,$$

where $\alpha_{ij}^{M,N}$ and β_i^M are appropriate constants obtained from the Gauss-Lobatto integration formulae. Thus obtaining a solution to (5.3.5) is equivalent to solving a least-squares problem. It has been shown in chapter II that if we choose our approximate solution $v^{M,N}$ such that it minimizes the modified functional

$$\tilde{\mathscr{X}}^{M,N}(w^{M,N}) = \sum_{j=0}^{2N} \sum_{i=0}^{4M-1} \alpha_{ij}^{M,N} |(L \ w^{M,N} - \bar{F}^{2M-1,2N-1})(\chi_i^{2M},\tau_j^{2N}) \ |^2$$

$$+ \sum_{i=0}^{4M-1} \beta_i^M \ | \ w^{M,N}(\chi_i^{2M},-1) - \bar{f}^{2M-1}(\chi_i^{2M})|^2$$

then we would be committing ,in addition , only a spectrally small further error. There is therefore no need to filter the coeffecients A and B in practice. We are interested in another aspect of this minimization procedure. Our approximate solution $v^{M,N}$ is the unique polynomial belonging to $(V^{M,N})^p$ which satisfies

$$(5.3.7) \qquad \int\int_{\Omega \times J} (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1})^* (L^{M,N} y^{M,N}) \, dx dt$$

$$\int_{\Omega \times (-1)} (v^{M,N} - \bar{f}^{2M-1})^* y^{M,N} \, dx = 0 \; ,$$

for all $y^{M,N} \in (V^{M,N})^p$.

We shall now use the above relation to prove that

$$\| L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega \times J} \leq C_s N^{1-s} \; ,$$

and

$$\| v^{M,N} - \bar{f}^{2M-1} \|_{-s,\Omega \times \{-1\}} \leq C_s N^{1-s} \; ,$$

for any $s > 1$.

In addition to this we shall also prove

$$\| L^{M,N} u - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega \times J} \leq C_s N^{-s}.$$

and

$$\| u - \bar{f}^{2M-1} \|_{-s,\Omega \times \{-1\}} \leq C N^{-s}.$$

With these results established we can prove Theorem 5.3.1 and the reader is advised to proceed directly to the theorem on page - 77 and continue his perusal of how these result are established only afterwards.

We first need to establish an upper bound on $\mathscr{X}^{M,N}(v^{M,N})$.

Let $w^{M,N}(x,t) = 0$. Then

$$(5.3.8) \qquad \mathscr{X}^{M,N}(w^{M,N}) \leq \| \bar{F}^{2M-1,2N-1} \|^2_{0,\Omega \times J} + \| \bar{f}^{2M-1} \|^2_{\Omega \times \{-1\}}$$

$$\leq \| F \|^2_{0,\Omega \times J} + \| f \|^2_{\Omega \times \{-1\}}$$

using (5.3.2). Hence we can conclude that

$$(5.3.9) \qquad \mathscr{X}^{M,N}(v^{M,N}) \leq C.$$

To estimate $\| L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega \times J}$ we need to bound

$$\frac{\left| (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, \phi)_{\Omega \times J} \right|}{\| \phi \|_{s,\Omega \times J}} \quad , \quad \text{for } \phi \in H.$$

Consider the periodic IVP

(5.3.10a) $\qquad L^{M,N} \psi = \phi \qquad$ for $(x,t) \in \Omega \times J$,

(5.3.10b) $\qquad\qquad \psi = 0 \qquad$ for $(x,t) \in \Omega \times \{-1\}$.

Then ψ is a smooth function and using estimate (5.2.3) we have

(5.3.11) $\qquad \| \psi \|_{s,\Omega \times J} \leq C_s \| \phi \|_{s,\Omega \times J}$,

where C_s is a constant which depends only on the smoothness of the coeffecients of the modified IVP and hence of the original IVP.

Let $Q^{M,N}$ be the projection operator that maps functions belonging to $H \cap H_{1,\Omega \times J}$ into $V^{M,N}$ defined as:

$Q^{M,N}w$ is the unique element of $V^{M,N}$ such that

$$\| w - Q^{M,N}w \|_{1,\Omega \times J} = \inf_{s^{M,N} \in V^{M,N}} \| w - s^{M,N} \|_{1,\Omega \times J} .$$

Then it is known that

(5.3.12) $\qquad \| w - Q^{M,N}w \|_{1,\Omega \times J} \leq C N^{1-s} \| w \|_{s,\Omega \times J}.$

Let $\tilde{\psi}^{M,N} = Q^{M,N} \psi$. Now

$$(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, \phi)_{\Omega \times J}$$

$$= (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N} \psi)_{\Omega \times J}$$

$$= (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N} \tilde{\psi}^{M,N})_{\Omega \times J}$$

$$+ (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N}(\psi - \tilde{\psi}^{M,N}))_{\Omega \times J} .$$

But by (5.3.7)

$$(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N} \tilde{\psi}^{M,N})_{\Omega \times J}$$

$$+ (v^{M,N} - \bar{f}^{2M-1}, \tilde{\psi}^{M,N})_{\Omega \times \{-1\}} = 0.$$

Now since $\psi = 0$ for $(x,t) \in \Omega \times \{-1\}$ we may write

$$(v^{M,N} - \bar{f}^{2M-1}, \tilde{\psi}^{M,N})_{\Omega \times \{-1\}}$$

$$= (v^{M,N} - \bar{f}^{2M-1}, \tilde{\psi}^{M,N} - \psi)_{\Omega \times \{-1\}}.$$

Hence we can conclude that

$$(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, \phi)_{\Omega \times J}$$

$$(5.3.13) \qquad = (v^{M,N} - \bar{f}^{2M-1}, \psi - \tilde{\psi}^{M,N})_{\Omega \times \{-1\}}$$

$$+ (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N}(\psi - \tilde{\psi}^{M,N}))_{\Omega \times J}.$$

Now using (5.3.12) we can conclude that

$$\| L^{M,N}(\psi - \tilde{\psi}^{M,N}) \|_{0,\Omega \times J} \le C N^{1-s} \| \psi \|_{s,\Omega \times J}.$$

And applying (5.3.11) we may write

$$(5.3.14) \qquad \| L^{M,N}(\psi - \tilde{\psi}^{M,N}) \|_{0,\Omega \times J} \le C N^{1-s} \| \phi \|_{s,\Omega \times J}.$$

But

$$(5.3.15) \qquad |(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1}, L^{M,N}(\psi - \tilde{\psi}^{M,N}))_{\Omega \times J}|$$

$$\le \| L^{M,N}(\psi - \tilde{\psi}^{M,N}) \|_{0,\Omega \times J} \times \| L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} \|_{0,\Omega \times J}$$

$$\le C_s N^{1-s} \| \phi \|_{s,\Omega \times J},$$

using (5.3.9) and (5.3.14).

Next, we estimate

$$|(v^{M,N} - \bar{f}^{2M-1}, \tilde{\psi}^{M,N} - \psi)_{\Omega \times \{-1\}}|.$$

From (5.3.9) we have that

(5.3.16)
$$\| v^{M,N} - \bar{f}^{2M-1} \|_{0,\Omega\times\{-1\}} \leq C.$$

Now

$$\| \tilde{\psi}^{M,N} - \psi \|_{0,\Omega\times\{-1\}} \leq C \| \tilde{\psi}^{M,N} - \psi \|_{1,\Omega\times J} \,,$$

by the trace theorem, and so by (5.3.12) we obtain

$$\| \tilde{\psi}^{M,N} - \psi \|_{0,\Omega\times\{-1\}} \leq C N^{1-s} \| \psi \|_{s,\Omega\times J}.$$

Using estimate (5.3.11) once again we conclude that

(5.3.17)
$$\| \tilde{\psi}^{M,N} - \psi \|_{0,\Omega\times\{-1\}} \leq C N^{1-s} \| \phi \|_{s,\Omega\times J}.$$

Hence applying (5.3.16) and (5.3.17) we get

(5.3.18)
$$|(v^{M,N} - \bar{f}^{2M-1} , \tilde{\psi}^{M,N} - \psi)_{\Omega\times\{-1\}}|$$

$$\leq C_s N^{1-s} \| \phi \|_{s,\Omega\times J}.$$

Combining (5.3.13) , (5.3.15) and (5.3.18) we obtain

$$|(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} , \phi)_{\Omega\times J}| \leq C_s N^{1-s} \| \phi \|_{s,\Omega\times J} \,,$$

and this gives us the required estimate

(5.3.19)
$$\| L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega\times J} \leq C_s N^{1-s}.$$

Next, we estimate

$$\| v^{M,N} - \bar{f}^{2M-1} \|_{-s,\Omega\times\{-1\}}$$

Consider the periodic IVP

(5.3.20a)
$$L^{M,N} \psi = 0 \qquad \text{for } (x,t) \in \Omega\times J \,,$$

(5.3.20b)
$$\psi = \mu \qquad \text{for } (x,t) \in \Omega\times\{-1\}.$$

Then ψ is a smooth function and using estimate (5.2.3) we have

(5.3.21)
$$\| \psi \|_{s,\Omega\times J} \leq C_s \| \mu \|_{s,\Omega\times\{-1\}}.$$

Let $\tilde{\psi}^{M,N} = Q^{M,N} \psi$. Now

$$(v^{M,N} - \bar{f}^{2M-1} , \mu)_{\Omega \times \{-1\}} = (v^{M,N} - \bar{f}^{2M-1} , \tilde{\psi}^{M,N})_{\Omega \times \{-1\}}$$

$$+ (v^{M,N} - \bar{f}^{2M-1} , \psi - \tilde{\psi}^{M,N})_{\Omega \times \{-1\}} .$$

But by (5.3.7)

$$(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} , L^{M,N} \tilde{\psi}^{M,N})_{\Omega \times J}$$

$$+ (v^{M,N} - \bar{f}^{2M-1} , \tilde{\psi}^{M,N})_{\Omega \times \{-1\}} = 0.$$

And since $L^{M,N} \psi = 0$ for $(x,t) \in \Omega \times J$ we may write

$$(L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} , L^{M,N} \tilde{\psi}^{M,N})_{\Omega \times J}$$

$$= (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} , L^{M,N} (\tilde{\psi}^{M,N} - \psi))_{\Omega \times J} .$$

Hence we can conclude that

$$(v^{M,N} - \bar{f}^{2M-1} , \mu)_{\Omega \times \{-1\}}$$

$$(5.3.22) \quad = (v^{M,N} - \bar{f}^{2M-1} , \psi - \tilde{\psi}^{M,N})_{\Omega \times \{-1\}}$$

$$+ (L^{M,N} v^{M,N} - \bar{F}^{2M-1,2N-1} , L^{M,N}(\psi - \tilde{\psi}^{M,N}))_{\Omega \times J} .$$

Thus we can show

$$(5.3.23) \quad \| v^{M,N} - \bar{f}^{2M-1} \|_{-s,\Omega \times \{-1\}} \le C_s N^{1-s} ,$$

using (5.3.22) and the arguments employed earlier.

We now need to estimate

$$\| L^{M,N} u - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega \times J} .$$

We know that u satisfies $u_t - A u_x - B u = F$ in the sense of distributions. Accordingly we may write

$$(5.3.24) \quad L^{M,N} u - \bar{F}^{2M-1,2N-1} = (L^{M,N} u - Lu) - (\bar{F}^{2M-1,2N-1} - F)$$

$$= -(\bar{A}^{M-1,N-1} - A) u_x - (\bar{B}^{M-1,N-1} - B) u - (\bar{F}^{2M-1,2N-1} - F).$$

Now by (5.3.4)

$$(5.3.25) \quad \| A - \bar{A}^{M-1,N-1} \|_{s,\infty,\Omega \times J} \le C_s N^{-s} \| A \|_{3s,\infty,\Omega \times J} ,$$

and so

$$(5.3.26) \qquad \| \bar{A}^{M-1,N-1} \|_{s,\infty,\Omega\times J} \leq C \| A \|_{3s,\infty,\Omega\times J} \quad ,$$

for M and N large enough.

Let us show how to estimate the various terms in (5.3.24). It is known (Canuto *et.al.* 1988) that the projection operator has the property that

$$(5.3.27) \qquad \| F - \bar{F}^{2M-1,2N-1} \|_{-s,\Omega\times J} \leq C N^{-s} \| F \|_{0,\Omega\times J} \quad .$$

Next , we shall estimate $\| (B - \bar{B}^{M-1,N-1}) u \|_{-s,\Omega\times J}$.

For this we need a lemma.

Lemma 5.3.1 Let $A \in H_{s,\infty,\Omega\times J}$ and $v \in H_{-s,\Omega\times J}$. Then $Av \in H_{-s,\Omega\times J}$ and

$$(5.3.28) \qquad \| Av \|_{-s,\Omega\times J} \leq C_s \| A \|_{s,\infty,\Omega\times J} \times \| v \|_{-s,\Omega\times J} \quad .$$

We have

$$(Av, \phi)_{\Omega\times J} = (v, A^*\phi)_{\Omega\times J} \quad , \text{ by definition.}$$

Hence

$$\frac{|(Av, \phi)_{\Omega\times J}|}{\| \phi \|_{s,\Omega\times J}} = \frac{|(v, A^*\phi)_{\Omega\times J}|}{\| \phi \|_{s,\Omega\times J}}$$

$$= \frac{|(v, A^*\phi)_{\Omega\times J}|}{\| A^*\phi \|_{s,\Omega\times J}} \times \frac{\| A^*\phi \|_{s,\Omega\times J}}{\| \phi \|_{s,\Omega\times J}} \quad .$$

And this gives

$$\| Av \|_{(-s),\Omega\times J} \leq \sup_{\phi \in H} \frac{\| A^*\phi \|_{s,\Omega\times J}}{\| \phi \|_{s,\Omega\times J}} \times \| v \|_{-s,\Omega\times J}$$

Now it is easy to see that

$$\sup_{\phi \in H} \frac{\| A^* \phi \|_{s,\Omega \times J}}{\| \phi \|_{s,\Omega \times J}} \leq C_s \| A \|_{s,\infty,\Omega \times J} .$$

And this gives us the required result.

Thus we obtain

$$\|(B - \bar{B}^{M-1,N-1}) u\|_{-s,\Omega \times J} \leq$$

$$C_s \|B - \bar{B}^{M-1,N-1}\|_{s,\infty,\Omega \times J} \times \| u \|_{-s,\Omega \times J} .$$

But

$$\| u \|_{-s,\Omega \times J} \leq \| u \|_{0,\Omega \times J} ,$$

and

$$\| B - \bar{B}^{M-1,N-1} \|_{s,\infty,\Omega \times J} \leq C_s N^{-s} \| B \|_{3s,\infty,\Omega \times J} .$$

Hence we obtain

(5.3.29) $$\| (B - \bar{B}^{M-1,N-1}) u \|_{s,\infty,\Omega \times J} \leq C_s N^{-s} .$$

Next, we estimate $\| u_x \|_{-s,\Omega \times J}$.

Let $\phi \in H$. Then

$$(u_x, \phi)_{\Omega \times J} = - (u, \phi_x)_{\Omega \times J} ,$$

since both u and ϕ are periodic in x.

Hence

$$\frac{|(u_x, \phi)_{\Omega \times J}|}{\| \phi \|_{s,\Omega \times J}} = \frac{|(u, \phi_x)_{\Omega \times J}|}{\| \phi \|_{s,\Omega \times J}}$$

But

$$\| \phi_x \|_{(s-1),\Omega \times J} \leq \| \phi \|_{s,\Omega \times J} .$$

And so we can conclude that

$$\sup_{\phi \in H} \frac{\left|(u_x, \phi)_{\Omega \times J}\right|}{\|\phi\|_{s, \Omega \times J}} \leq \sup_{\phi \in H} \frac{\left|(u, \phi_x)_{\Omega \times J}\right|}{\|\phi_x\|_{s-1, \Omega \times J}},$$

which gives us

(5.3.30) $\qquad \|u_x\|_{-s, \Omega \times J} \leq \|u\|_{-s+1, \Omega \times J}$.

But

(5.3.31) $\qquad \|u\|_{-s+1, \Omega \times J} \leq \|u\|_{0, \Omega \times J}$.

And so by the lemma just proved we get

(5.3.32) $\qquad \|(A - \overline{A}^{M-1, N-1}) u_x\|_{s, \infty, \Omega \times J} \leq C_s N^{-s}$.

Combining all these estimates we get the required result

(5.3.33) $\qquad \|L^{M,N} u - \overline{F}^{2M-1, 2N-1}\|_{-s, \Omega \times J} \leq C_s N^{-s}$.

Also we have (Canuto *et.al.* 1988)

(5.3.34) $\qquad \|u - \overline{f}^{2M-1}\|_{-s, \Omega \times \{-1\}} \leq C N^{-s}$.

We can now prove our main theorem.

Theorem 5.3.1 Let $v^{M,N}$ be the solution obtained by minimizing $\mathscr{H}^{M,N}(w^{M,N})$ as described in (5.3.5). Then for all $s \geq 0$ the estimate

(5.3.35) $\qquad \|u - v^{M,N}\|_{-s, \Omega \times \{1\}} + \|u - v^{M,N}\|_{-s, \Omega \times J} \leq C_s N^{1-s}$

holds.

We have by (5.3.19) that

$$\|L^{M,N} v^{M,N} - \overline{F}^{2M-1, 2N-1}\|_{-s, \Omega \times J} \leq C_s N^{1-s}.$$

Moreover, by (5.3.33) we know that

$$\|L^{M,N} u - \overline{F}^{2M-1, 2N-1}\|_{-s, \Omega \times J} \leq C_s N^{1-s}.$$

Using the triangle inequality we obtain

$$(5.3.36) \qquad \| L^{M,N} (u - v^{M,N}) \|_{-s, \Omega \times J} \le C_s N^{1-s}.$$

Finally , we have

$$(5.3.37) \qquad \| u - v^{M,N} \|_{-s, \Omega \times \{-1\}}.$$

$$\le \| u - \bar{f}^{2M-1} \|_{-s, \Omega \times \{-1\}} + \| \bar{f}^{2M-1} - v^{M,N} \|_{-s, \Omega \times \{-1\}}$$

$$\le C_s N^{1-s} ,$$

using (5.3.23) and (5.3.34).

Therefore using estimate (5.2.4) along with (5.3.36) and (5.3.37) we conclude that

$$\| u - v^{M,N} \|_{-s, \Omega \times \{1\}} + \| u - v^{M,N} \|_{-s, \Omega \times J}$$

$$\le C_s N^{1-s} . \qquad \blacksquare$$

5.4   Recovering Pointwise Values with Spectral Accuracy

In this section we briefly describe how the local smoothing proposed by Gottlieb, Tadmor(1985) and Abarbanel, Gottlieb(1985) and can be used to recover pointwise values with spectral accuracy at any point in a neighbourhood of which the actual solution is smooth. If we wish to recover the values at t = 1 the local smoothing is particularly simple. Suppose we wish to obtain the value of the solution at the point $(x_0, 1)$. We assume that there exists a neighbourhood

$$J = \{ x : | x - x_0 | \le \delta \}$$

in which the actual solution u(x,t) is smooth. Let $\rho(x)$ be a C_0^∞ function with support in the set J and such that ρ is nonnegative everywhere and $\rho(x_0) = 1$. Choose $K = M^\beta$ with $0 < \beta < 1$, and let $D^K(\xi)$ denote the Dirichlet kernel

$$D^K(\xi) = \sum_{J=-K}^{K} e^{1J\xi} = \frac{\sin((2K+1)\xi/2)}{\sin(\xi/2)} \quad , \; \xi \neq 2m\pi \; ,$$

$$= 2K+1 \qquad\qquad , \; \xi = 2m\pi.$$

Then to obtain the regularized version of $v^{M,N}$ at $(x_0,1)$ we define

$$(5.4.1) \qquad R \, v^{M,N}(x_0,1) = \frac{1}{2\pi} \int_0^{2\pi} D^K(x_0 - x) \, \rho(x) \, v^{M,N}(x,1) \, dx.$$

It has been proved in (Canuto *et.al.* 1988) that if

$$\| u - v^{M,N} \|_{-s,\Omega\times\{1\}} \leq C_s \, M^{-s+1} \; , \; \text{then}$$

$$(5.4.2) \qquad |u(x_0,1) - R \, v^{M,N}(x_0,1)| \leq C_1 \, (1 + \log M) \, M^{-s+1}$$
$$+ \; C_2 \, M^{-s+1+\beta s} \; ,$$

where the constants C_1 and C_2 depend upon the Sobolev norms of ρ and u over the interval J. A balance of the errors is acheived by putting $\beta = 1/2$, in which case we obtain

$$|u(x_0,1) - R \, v^{M,N}(x_0,1)| = O \, (M^{-s/2+1}) \; ,$$

which proves that $u(x_0,1)$ can be approximated with spectral accuracy starting from the knowledge of the Galerkin-Collocation approximation $v^{M,N}$.

Suppose now that we wish to recover the value of the solution at an interior point (x_0, t_0). We assume that $u(x,t)$ is smooth in the set O , where

$$O = \{ (x,t) : |x - x_0| \leq \delta \; , \; |t - t_0| \leq \epsilon \}.$$

Let $\rho(x)$ be a C_0^∞ function with support in the set

$$J = \{ x : |x - x_0| < \delta \} \; ,$$

which is nonnegative everywhere and such that $\rho(x_0)$

Similarily let $\eta(t)$ be a C_0^∞ function with support in the set

$$K = \{ t : |t - t_0| < \epsilon \} ,$$

which is nonnegative everywhere and satisfying $\eta(t_0) = 1$. Choose $K = M^\beta$ and $L = N^\gamma$ with $0 < \beta$, $\gamma < 1/2$. Let $D^K(\xi)$ denote the Dirichlet kernel and $E^L(\tau, \tau_0)$ denote the Legendre kernel

$$E^L(\tau, \tau_0) = \sum_{j=0}^{L} (j + 1/2) \, L_j(\tau) \, L_j(\tau_0).$$

Then to obtain the regularized values of $v^{M,N}$ at (x_0, t_0) we define

$$R \, v^{M,N}(x_0, t) = \frac{1}{2\pi} \int\int_{Q \times J} D^K(x_0 - x) \, E^L(t, t_0) \, \rho(x) \, \eta(t) \, v^{M,N}(x,t) dx dt.$$

Once more it can be shown that $R \, v^{M,N}(x_0, t_0)$ approximates $u(x_0, t_0)$ with spectral accuracy and an optimal balance of the errors is obtained by choosing $\beta = \gamma = 1/3$.

5.5 Numerical Results

In this section we demonstrate the efficacy of the method proposed in this paper.

Example 5.5.1

Consider the initial value problem with periodic boundary .

$$U_t - a(x,t) \, U_x - b(x,t) \, U = F(x,t)$$

$$0 \leq x \leq 2\pi, \; -1 \leq t \leq 1$$

with initial condition

$$U(x,-1) = g(x)$$

and let $g(x)$ has a discontinuty in its derivative .

Case I

Consider

$$U(x,t) = \begin{cases} (1 + t)t + \sin(x) & 0 \leq x < \pi \\ (1 + t)t - \sin(x) & \pi \leq x < 2\pi \end{cases}.$$

and take $a(x,t) = 0.5$ and $b(x,t) = 0.0$.

Case II

Consider

$$U(x,t) = \begin{cases} (1 + t)\sin(t) + x & 0 \leq x < \pi \\ (1 + t)\sin(t) + 2\pi - x & \pi \leq x < 2\pi \end{cases}.$$

and with same $a(x,t)$ and $b(x,t)$.

Results of smoothing of the spectral approximation of $U(x,t)$, with $M = 128$ and $N = 17$ are displayed below

Case I

$x_\nu = \frac{\pi}{8} (\nu+1/2)$ ν equals	$\|U(x_\nu,1) - V^{m,n}(x_\nu,1)\|$	$\|U(x_\nu,1) - R\ V^{m,n}(x_\nu,1)\|$
4	1.47 (−3)	2.69 (−8)
5	1.88 (−3)	2.28 (−8)
6	2.32 (−3)	2.67 (−8)

Table 5.5.1

Case II

$x_\nu = \frac{\pi}{8} (\nu+1/2)$ ν equals	$\|U(x_\nu,1) - V^{m,n}(x_\nu,1)\|$	$\|U(x_\nu,1) - R\ V^{m,n}(x_\nu,1)\|$
4	1.12 (−3)	2.88 (−8)
5	1.11 (−3)	4.19 (−8)
6	1.28 (−3)	4.48 (−8)

Table 5.5.2

where the number in bracket denote the power to base ten.

REFERENCES

[1]   ABARBANEL.S. , GOTTLIEB.D. & TADMOR.E. (1986): Spectral Methods for Discontinuous Problems, In Numerical Methods for Fluid Dynamics, ed. by K.W.Morton, M.J.Baines (Oxford University Press , London), pp 129.

[2]   CANUTO.C., HUSSAINI.M.Y., QUARTERONI.A. AND ZANG.T.A.(1987): Spectral Methods in Fluid Dynamics" Springer Series in Computational Physics, Springer Verlag .

[3]   DUTT.P. (1990): Spectral Methods For Initial Boundary Value Problems an Alternative Apporach. SIAM J.Numer.Anal., 27(4), pp 885.

[4]   FINLAYSON.B.A., SCRIEN.L.E. (1966): The Method of Weighted Residuals - a review. Appl.Mech.Rev. 19, pp 735.

[5]   GOLDBERG.M. AND TADMOR.E. (1980) : Convenient Stability Criteria for Difference Approximations of Hyperbolic initial - Boundary Value Problems Math Comput , 32,pp 885.

[6]   GOTTLIEB.D. ,HUSSAINI.M.Y. AND ORSZAG.S.A (1984): Theory and Application of spectral Methods, in Spectral Methods for Partial Differential Equations,Ed by R.G.Voigt , DGottlieb, M.Y.Hussaini ,SIAM —CBMS, pp 1.

[7]   GOTTLIEB.D. & TADMOR.E. (1985): Recovering Pointwise Values of Discontinuous Data Within Spectral Accuracy , In Progress and Supercomputing in Computational Fluid Dynamics , ed. by E.E.Murman & S.S.Abarbanel (Birkhauser, Boston), pp 357.

[8]   MAJDA.A., Mc DONNOUGH.J. & S. OSHER (1978): The Fourier Method for Nonsmooth Initial Data, Math. Comput. 32, pp 1041.

[9]   MERCIER.B. (1981): Analyse Numerique des Methodes Spectrales , Note CEA-N-2278 (Commissariat a l'Energie Atomique Centre d'Etudes de Limeil , 94190 Villeneuve - Saint Georges)

[10] MORCHOISNE.Y.(1979): Rsolution of Navier-Stokes equations by a Space-Time Pseudospectral MEthod. Rech.Aerosp., 5, pp 293.

[11] MORCHOISNE.Y. (1984): Inhomogenous Flow Calculations for Spectral Methods: Mono-Domain and Multi-Domain Techniques, in Spectral Methods for Partial Differential Equations, Ed by R.G.Voigt, D.Gottlieb, M.Y.Hussaini, SIAM -CBMS,pp 181.

[12] ORSZAG.A.S. (1980): Spectral Methods for Problems in Complex Geometries J.Comput.Phys ., 37, pp 70.

[13] PULLIAM.H.T. AND CHAUSEE.J. (1981): A Diagonal From of an Implicit Approximate-Factorization Algorithm., J.Comput.Phy, 39, pp 347.

[14] RAUCH.J. (1972): $\mathcal{L}_2$ is a Continuable Initial Condition for Kreiss Mixed Problems, Comm. Pure Appl. Math., 25, pp 265.

[15] STEGER.J.L. AND BUNING.P.G. (1985) : Developments in the Simulation of Compressible Inocid and viscous Flow on Supercomputers, in Progress and Super-Computing in Computational Fluid Dynamics, Ed by E.E.Murman, S.S.Abarbanel and Birkhäuser, Boston, pp 357.

[16] Szegö.G. (1939): Orthogonal Polynomials. Vol 23,(AMS Coll. Pubh. New York)

[17]   TAYLOR.M. (1981): Psudodifferential Operators (Princeton University Press ,NJ) .

[18] ZANG.T.A. ,WONG.Y.S. AND HUSSAINI.M.Y. (1984): Spectral
Multigrid,   Methods   for   Elliptic   Equations   II.,
J.Comput.Phys., 54, pp 489 .

)ATH-1992-D-S/N-SPE